

Routledge Encyclopedia of Philosophy

Logic & Mathematics

Contents

Adverbs	1
Ambiguity	6
Analysis, nonstandard	10
Analysis, philosophical issues in	16
Analyticity	21
Antirealism in the philosophy of mathematics	26
Arithmetic, philosophical issues in	30
Axiom of choice	35
Beth's theorem and Craig's theorem	36
Boolean algebra	43
Cantor's theorem	47
Category theory, applications to the foundations of mathematics	49
Category theory, introduction to	54
Chaos theory	57
Church's theorem and the decision problem	61
Church's thesis	63
Combinatory logic	68
Common-sense reasoning, theories of	71
Complexity, computational	74
Compositionality	79
Computability and information	81
Computability theory	87
Computer science	97
Consequence, conceptions of	103
Constructivism in mathematics	108
Continuum hypothesis	116
Conventionalism	120
Counterfactual conditionals	124
De re/de dicto	129
Decision and game theory	133
Definition	150
Demonstratives and indexicals	155
Deontic logic	160
Descriptions	164
Dialogical logic	170
Dynamic logics	173
Epistemic logic	176
Existence	183
Fallacies	186
Forcing	188
Formal and informal logic	193
Formal languages and systems	195

Contents

Free logics, philosophical issues in	199
Fuzzy logic	203
Geometry, philosophical issues in	207
Godel's theorems	214
Herbrand's theorem	228
Hilbert's programme and formalism	231
Identity	239
Identity of indiscernibles	242
Imperative logic	245
Indicative conditionals	249
Indirect discourse	252
Inductive definitions and proofs	255
Inference, Indian theories of	260
Infinitary logics	269
Infinity	272
Intensional entities	279
Intensional logics	284
Intensionality	288
Intuitionism	290
Intuitionistic logic and antirealism	298
Lambda calculus	306
Logic, ancient	311
Linear Logic	315
Logic in China	323
Logic in Islamic philosophy	337
Logic in Japan	345
Logic in the 17th and 18th centuries	349
Logic in the 19th century	356
Logic in the early 20th century	366
Logic machines and diagrams	376
Logic, medieval	388
Logic, Philosophy of	401
Logic, Renaissance	404
Logical constants	411
Logical form	417
Logical laws	423
Logical positivism	428
Logicism	434
Lowenheim-Skolem theorems and non-standard models	441
Many-valued logics	447
Many-valued logics, philosophical issues in	451
Mass terms	459

Contents

Mereology	462
Modal logic	465
Modal logic, philosophical issues in	475
Modal operators	483
Model theory	486
Models	494
Multiple-conclusion logic	499
Natural deduction, tableau and sequent systems	502
Necessary truth and convention	510
Non-constructive rules of inference	514
Non-monotonic logic	518
Numbers	523
Ontological commitment	530
Ordinal logics	535
Paraconsistent logic	540
Paradoxes of set and property	543
Polish logic	551
Predicate Calculus	560
Predication	570
Presupposition	572
Probability, interpretations of	576
Proof theory	580
Proper names	591
Property theory	597
Propositional attitude statements	604
Propositions, sentences and statements	614
Provability logic	616
Pythagoreanism	621
Quantification and inference	628
Quantifiers	631
Quantifiers, generalized	635
Quantifiers, substitutional and objectual	641
Quantum logic	646
Questions	651
Randomness	654
Realism in the philosophy of mathematics	659
Recursion-theoretic hierarchies	664
Reference	668
Relevance logic and entailment	681
Rhetoric	686
Scope	691
Second- and higher-order logics	693

Contents

Second-order logic, philosophical issues in	698
Semantic paradoxes and theories of truth	703
Semantics	711
Sense and reference	715
Set theory	720
Set theory, different systems of	731
Tarski's definition of truth	740
Tense and temporal logic	751
The constructible universe	757
Theory of types	762
Turing machines	766
Turing reducibility and Turing degrees	771
Type/token distinction	775
Use/mention distinction and quotation	777
Vagueness	781

Adverbs

Adverbs are so named from their role in modifying verbs and other non-nominal expressions. For example, in 'John ran slowly', the adverb 'slowly' modifies 'ran' by characterizing the manner of John's running. The debate on the semantic contribution of adverbs centres on two approaches. On the first approach, adverbs are understood as predicate operators: for example, in 'John ran slowly', 'ran' would be taken to be a predicate and 'slowly' an operator affecting its meaning. Working this out in detail requires the resources of higher-order logic. On the second approach, adverbs are understood as predicates of 'objects' such as events and states, reference to which is revealed in logical form. For example, 'John ran slowly' would be construed along the lines of 'there was a running by John and it was slow', in which the adverb 'slowly' has become a predicate 'slow' applied to the event that was John's running.

Since adverbs are exclusively modifiers, they are classed among the syncategorematic words of terminist logic, the investigation of which carried the subject forward from Aristotle in the thirteenth century. (The contrasting 'categoremata' - grammatical subjects and predicates - are those words which have meaning independently.) They are of contemporary interest for philosophical logic and semantic theory, because particular accounts of them carry implications for the nature of combinatorial semantics and language understanding, and for ontology.

1 Syntactic types and semantic combination

There are several types of adverbial constructions, of which we distinguish the following classes: (a) 'manner' adverbs, which intuitively function as simple modifiers of verbs; (b) 'thematic' adverbs, of which some and possibly all function as (at least) two-place predicates in their own right; (c) adverbs of quantification, which express generality applying to whole sentences; and (d) discourse particles, whose meaning evidently derives from their role in linking clauses or independent sentences. (These categories are not exhaustive.) 'Adverbs', especially manner 'adverbs', are not in fact confined to single words. The general category is therefore not that of adverbs, but of adverbial phrases or adverbials (for example, 'more quickly than Mary', 'very frequently').

Typical manner adverbials are as in (1) below, thematic adverbs as in (2) and adverbs of quantification as in (3):

- (1) John walked slowly/quietly/more quickly than Mary.
- (2) Mary apparently/reluctantly went to New York.
- (3) Mary occasionally/always walks to work.

Discourse particles, considered briefly below, include 'but', 'anyway' and several others. We discuss these cases in turn.

The essential logical problem of manner adverbials is already apparent in the simplest examples. A verb combines with a manner adverb to form a complex verbal construction of the same type. Thus 'walk' and 'walk slowly' are both predicates, and the syntax of the combination may be depicted as follows:

$[_V [_V \text{ walk}] [_{Adv} \text{ slowly}]]$

If (disregarding tense) we take 'walk' as a one-place predicate, then the semantics of this combination might be given by positing that 'slowly' is interpreted as a predicate operator; that is, as a function that maps one-place predicate interpretations onto other one-place predicate interpretations. Alternatively, it may be suggested that 'slowly' and the other manner adverbials are, logically speaking, predicates in their own right, specifically predicates of actions. The adjectives to which they are related do seem to play this role. Corresponding to (1), for instance, we have the adjectival predications

- (4) John's walk was slow/quiet/quicker than Mary's.

If we take the further step of supposing that the verb 'walk' is in fact a two-place predicate, with a position for actions, then the combination 'walk slowly' can be interpreted as

$\text{walk}(x, e) \ \& \ \text{slow}(e)$,

where e ranges over actions. Comparing this account with the first alternative, (slowly(walk))(x),

we see a trade-off: where predicates are taken to have a simple structure, the adverbial must be understood as an operator; but where extra structure, in the form of a place for actions, is posited, the semantic combination is truth-functional and predicate operators are not required.

The alternatives just sketched each have their defenders in the literature on adverbials. Adverbs are construed as predicate operators in formal theories of linguistic structure, including those of Montague (1974) and Lewis (1975). The predicative alternative was first advanced at length by Davidson (1967), and is elaborated by Parsons (1990). Semantic and metaphysical issues arise for each account; we take up some of these below.

Thematic adverbs are intuitively distinguished from manner adverbs in as much as they yield constructions adverb + verb which cannot be treated simply as new verbs: apparently going to New York is not a way of going to New York; and reluctantly going to New York is not a manner of travel, but an instance of travel whose agent was reluctant to undertake it. For the examples in (2) the following paraphrases suggest themselves:

It was apparent that Mary went to New York.

Mary went to New York and she was reluctant to go to New York.

The discourse particles, traditionally and appropriately called adverbs, have come under relatively formal study only in recent years. The following examples are representative:

(5) He was poor *but* honest.

(6) *Anyway*, I'm going to New York.

They resist analysis in terms of their contribution to truth-conditions, but carry implications for the discourses in which they occur, with (5) involving some presumptive contrast with what might have been expected and (6), as an assertion, functioning to indicate a return to a superordinate topic of conversation. See Levinson (1983) for examples and discussion.

2 Ambiguities

Many adverbs are ambiguous between thematic-adverbial and manner-adverbial interpretation. In an example such as

(7) Mary quickly objected,

we may have either the interpretation 'Mary's objection was delivered in a quick manner' or the interpretation 'Mary's objection came a short time after the enunciation of the proposition to which she objected'. The second, thematic-adverbial interpretation shows up in the corresponding adjective 'quick' in a construction such as

Mary was quick to object.

These examples suggest that grammatical appearance belies logical structure, since the thematic adverbial in (7) functions, logically speaking, as the main predicate of that sentence. Austin (1956) observed that adverbial position often disambiguates, with post-verb adverbs favouring the manner-interpretation, and pre-verb adverbs the thematic, as in the pair

He trod on the snail clumsily.

Clumsily, he trod on the snail.

Adverbs of quantification pick up arguments including, but not restricted to, the temporal. Sentences such as (8) are ambiguous, depending upon whether the quantification is over occasions, or over the subject:

(8) Travel books are seldom worth reading.

Where the adverb quantifies over time, (8) means that the occasions are few when it is worth reading travel books. But there is another salient interpretation:

Few travel books are worth reading.

Where quantification over time would be ridiculous in view of the subject matter, construal of the adverb with the subject is particularly salient, for example:

Quadratic equations seldom have real solutions.

The above reflections have it that thematic adverbs and adverbs of quantification are not modifiers at all, except in a purely grammatical sense. Inversely, there have been suggestions that what appear in language as if they were arguments to a predicate actually function as adverbial modifiers. Perhaps the best known account of this type is Roderick Chisholm's discussion of a certain class of statements about appearances. Chisholm suggests that a man who 'sees spots before his eyes' should be thought of as 'sensing in a spotty manner' or as someone whom things 'appeared to spottily' (1957). This philosophical move is designed to rid the locution of any implication that in sensing spots before his eyes the man is sensing a mysterious object, an appearance, which is before him, his eyes or his mind (see *Mental states, adverbial theory of*). In a similar vein, Goodman ([1968] 1976) construed the locution ' x represents an F ', under the condition where there is no implication that such an F exists, as 'is an F -representation', effectively treating the predicate F adverbially.

The suggestions of Chisholm and Goodman, left as they are, become problematic if the project of giving a combinatorial semantics for language is taken seriously (see *Compositionality*). Suppose x is a picture and (9) is true, with no existential implications:

(9) x represents two unicorns galloping in a field.

If we write, with Goodman,

x is a two-unicorns-galloping-in-a-field representation,

then we have yet to provide a semantic structure to go along with the syntax. But there must be some such structure, since, for example, (9) obviously implies

x represents more than one unicorn in motion.

Similar issues arise for the predicate-operator theory of manner adverbs. For instance, that theory does not immediately deliver the obvious implication

(10) John walked slowly; therefore, John walked.

In Montague (1974) and much subsequent literature, this and similar implications are the consequences of semantic postulates in the sense of Carnap (1956). Even with such postulates, the relation between the adverbial constructions in (1) and their adjectival paraphrases (4) requires clarification. By contrast, the predicative theory proposed by Davidson is specifically designed to capture such implications and paraphrases. The premise of (10) is rendered as

(11) $(\exists e)(\text{walk}(\text{John}, e) \ \& \ \text{slow}(e))$,

that is, 'There was a walk by John and it was slow'; and the conclusion has the form

(12) $(\exists e)\text{walk}(\text{John}, e)$,

'There was a walk by John', a trivial implicate of the premise. For the paraphrase relation between (1) and (4), we have only to note that the complex noun 'John's walk' would be understood as a definite description of an action; that is, as '(the e) walk(John , e)'. 'John's walk was slow' then becomes

(13) $\text{slow}((\text{the } e) \text{ walk}(\text{John}, e))$.

Given any standard treatment of the definite description, (13) will imply (11). However, the Davidsonian view is committed to supplying extra, hidden structure in all cases, and to taking events as individuals, a step that has often been considered metaphysically dubious.

Ensembles of adverbs show ambiguity of scope (see [Scope](#)). For example, one interpretation of (14) has John being clever in that he made a stupid response:

(14) John cleverly responded stupidly.

Modal adverbs such as ‘necessarily’ and ‘contingently’, which modify whole sentences, allow singular terms and quantifiers within their superficial scope to be interpreted as outside it (a point known to the terminist logicians). Tracking the relative scopes of modalities is part of the contemporary application of modal logic. As in the cases of manner and thematic adverbs, there are both operator-theories and predicate-theories of these expressions; that is, the modalities may be developed as one-place modal operators, with the same syntax as negation; or as predicates of sentences in a first-order formulation of the logical syntax of natural language (see [Modal operators](#)). Montague (1963) argued that the latter was unacceptable since the normal laws of modal logic could not all be maintained, on pain of paradox. The argument has subsequently been developed and discussed by others: see Koons (1992) for a survey and response.

3 Extensionality

Adverbial constructions of both the manner and thematic types, together with others, show a kind of superficial but persistent non-extensionality: superficial, because it may disappear under analysis; and persistent, because it may show up in places that the analysis itself uncovers. Predicate-operator theories can accept non-extensionality as the norm. Supposing, for example, that those who breathe are exactly those who perceive, it is absurd to infer that *x* perceives rapidly from the premise that *x* breathes rapidly. The predicate-operator theory, taking all operations to be in intension, has no such consequence. Or supposing that those who went to New York were exactly those who visited Times Square, it does not follow that if Mary reluctantly went to New York she also reluctantly visited Times Square.

For theories of Davidson’s type, the constructions

x breathed rapidly
x perceived rapidly

are of the same logical type, but the events on which the adverb is predicated are different. The coextensiveness of ‘breathe’ and ‘perceive’ amounts to the coextensiveness of

(for some *e*) breathe(*x*, *e*)
 (for some *e*) perceive(*x*, *e*).

Such coextensiveness no more implies the equivalence of ‘*x* breathed rapidly’ to ‘*x* perceived rapidly’ than the coextensiveness of ‘*x* kicked something’ and ‘*x* saw something’ would imply the equivalence of ‘*x* saw something red’ and ‘*x* kicked something red’.

Non-extensionality is more troublesome when one considers events related as genus and species. For example, ‘Mary flew slowly across Spain’ does not imply ‘Mary travelled slowly across Spain’, even though any event of flying is an event of travelling. It follows that it is inadequate to represent ‘Mary flew slowly across Spain’ simply as

(for some *e*) (fly(Mary, *e*) & across Spain(*e*) & slow(*e*)).

Rather, we must add that the event *e*, which was a flying and therefore a travelling, was slow *for a flying*. These examples show at least that manner adverbials are relative to the sets of events against which a given event is evaluated. Other examples may show that it is not merely sets but also properties, not extensionally individuated, that form the background.

In any case, non-extensionality is evident in constructions with thematic adverbs. Davidson, following a suggestion by Hector-Neri Castañeda, considered examples such as that of

Oedipus intentionally married Jocasta,

from which it does not follow that Oedipus intentionally married his mother. In this case, the locus of non-extensionality, although due to the presence of the adverb ‘intentionally’, is the predicate ‘intend’, from which the adverb is derived. In this sense, it is independent of any peculiarities of adverbs, as in the Davidsonian paraphrase

Oedipus married Jocasta, and he intended to marry Jocasta.

Higginbotham (1989) explores a number of similar examples.

See also: [Logical and mathematical terms, glossary of](#); [Syntax §6](#)

JAMES HIGGINBOTHAM

References and further reading

- Austin, J.L.** (1956) ‘A Plea for Excuses’, *Proceedings of the Aristotelian Society*; repr. in *Philosophical Papers*, Oxford: Clarendon Press, 1961, 123-52. (A famous and highly readable paper in which Austin tries to clarify the nature of action from the way in which we use excuses to escape responsibility for the consequences of what we do.)
- Carnap, R.** (1956) *Meaning and Necessity*, Chicago, IL: University of Chicago Press. (Develops the idea that many apparently logical inferences are to be accounted for by semantic postulates that relate the meanings of the words in premises and conclusion; quite technical.)
- Chisholm, R.** (1957) *Perceiving: A Philosophical Study*, Ithaca, NY: Cornell University Press, esp. ch. 8. (Proposes an adverbial analysis of perception statements to avoid commitment to such entities as ‘sense-data’.)
- Davidson, D.** (1967) ‘The Logical Form of Action Sentences’, in N. Rescher (ed.) *The Logic of Decision and Action*, Pittsburgh, PA: University of Pittsburgh Press; repr. in *Essays on Actions and Events*, Oxford: Clarendon Press, 1980, 105-48. (An analysis of adverbial modification as predication of events; readable and influential.)
- Goodman, N.** (1968) *Languages of Art*, Indianapolis, IN: Hackett Publishing Company, 2nd edn, 1976. (Proposes an adverbial analysis of the likes of ‘is a picture of an *F*’ to prevent the existence of a picture of an *F* entailing the existence of a pictured *F*.)
- Higginbotham, J.** (1989) ‘Elucidations of Meaning’, *Linguistics and Philosophy* 12: 465-517. (§2 defends a Davidsonian analysis in the context of a discussion of how lexical analysis interacts with permissible modes of syntactic composition; quite technical.)
- Koons, R.C.** (1992) *Paradoxes of Belief and Strategic Rationality*, Cambridge: Cambridge University Press. (Mainly concerns the role of common knowledge or mutual belief in generating liar-like paradoxes about rational action; very technical.)
- Kretzmann, N.** (1968) *William of Sherwood’s Treatise on Syncategorematic Words*, Minneapolis, MN: University of Minnesota Press. (Elucidates the categorematic/syncategorematic distinction mentioned in the introduction to this entry.)
- Levinson, S.C.** (1983) *Pragmatics*, Cambridge: Cambridge University Press. (Includes discussion of words such as ‘but’ and ‘anyway’ which are traditionally regarded as adverbs but whose contribution to meaning is primarily at the level of pragmatics.)
- Lewis, D.K.** (1975) ‘Adverbs of Quantification’, in E. Keenan (ed.) *Formal Semantics of Natural Language*, Cambridge: Cambridge University Press, 3-15. (Semantics for adverbs construed as predicate operators.)
- Montague, R.** (1963) ‘Syntactical Treatment of Modality, with Corollaries on Reflexion Principles and Finite Axiomatizability’, *Acta Philosophica Fennica* 16; repr. in *Formal Philosophy*, ed. R. Thomason, New Haven, CT: Yale University Press, 1974, 286-302. (Shows that straightforward analysis of modal adverbs (‘necessarily’, ‘possibly’) as predicates of sentences (‘is necessarily true’, ‘is possibly true’) leads to the generation of a paradox similar to Tarski’s in his proof of the ‘undefinability of truth’; very technical.)
- Montague, R.** (1974) *Formal Philosophy*, ed. R. Thomason, New Haven, CT: Yale University Press. (Montague’s collected papers in natural language semantics; extremely technical. For adverbs, see especially ‘English as a Formal Language’.)
- Parsons, T.** (1990) *Events in the Semantics of English*, Cambridge, MA: MIT Press. (Extended development and discussion of Davidson’s analysis of adverbs; clear and, for the most part, non-technical; good bibliography.)

Ambiguity

A word, phrase or sentence is ambiguous if it has more than one meaning. The word 'light', for example, can mean not very heavy or not very dark. Words like 'light', 'note', 'bear' and 'over' are lexically ambiguous. They induce ambiguity in phrases or sentences in which they occur, such as 'light suit' and 'The duchess can't bear children'. However, phrases and sentences can be ambiguous even if none of their constituents is. The phrase 'porcelain egg container' is structurally ambiguous, as is the sentence 'The police shot the rioters with guns'. Ambiguity can have both a lexical and a structural basis, as with sentences like 'I left her behind for you' and 'He saw her duck'.

The notion of ambiguity has philosophical applications. For example, identifying an ambiguity can aid in solving a philosophical problem. Suppose one wonders how two people can have the same idea, say of a unicorn. This can seem puzzling until one distinguishes 'idea' in the sense of a particular psychological occurrence, a mental representation, from 'idea' in the sense of an abstract, shareable concept. On the other hand, gratuitous claims of ambiguity can make for overly simple solutions. Accordingly, the question arises of how genuine ambiguities can be distinguished from spurious ones. Part of the answer consists in identifying phenomena with which ambiguity may be confused, such as vagueness, unclarity, inexplicitness and indexicality.

1 Types of ambiguity

Ambiguity is a property of linguistic expressions. A word, phrase or sentence is ambiguous if it has more than one meaning. Obviously this definition does not say what meanings are or what it is for an expression to have one (or more than one). For a particular language, this information is provided by a grammar, which systematically pairs forms with meanings, ambiguous forms with more than one meaning (see [Semantics](#)).

There are two types of ambiguity, *lexical* and *structural*. Lexical ambiguity is by far the more common. Everyday examples include nouns like 'chip', 'pen' and 'suit', verbs like 'call', 'draw' and 'run' and adjectives like 'deep', 'dry' and 'hard'. There are various tests for lexical ambiguity. One test is having two unrelated antonyms, as with 'hard', which has both 'soft' and 'easy' as opposites. Another is the conjunction reduction test. Consider the sentence, 'The tailor pressed one suit in his shop and one in the municipal court'. Evidence that the word 'suit' (not to mention 'press') is ambiguous is provided by the anomaly of the 'crossed interpretation' of the sentence, on which 'suit' is used to refer to an article of clothing and 'one' to a legal action.

The above examples of ambiguity are each a case of one word with more than one meaning. However, it is not always clear when we have only one word. The verb 'desert' and the noun 'dessert', which sound the same but are spelled differently, count as distinct words (they are homonyms). So do the noun 'bear' and the verb 'bear', even though they not only sound the same but are spelled the same. These examples may be clear cases of homonymy, but what about the noun 'respect' and the verb 'respect' or the preposition 'over' and the adjective 'over'? Are the members of these pairs homonyms or different forms of the same word? There is no general consensus on how to draw the line between cases of one ambiguous word and cases of two homonymous words. Perhaps the difference is ultimately arbitrary.

Sometimes one meaning of a word is derived from another. For example, the cognitive sense of 'see' (to see that something is so) seems derived from its visual sense. The sense of 'weigh' in 'He weighed the package' is derived from its sense in 'The package weighed two pounds'. Similarly, the transitive senses of 'burn', 'fly' and 'walk' are derived from their intransitive senses. Now it could be argued that in each of these cases the derived sense does not really qualify as a second meaning of the word but is actually the result of a lexical operation on the underived sense. This argument is plausible to the extent that the phenomenon is systematic and general, rather than peculiar to particular words. Lexical semantics has the task of identifying and characterizing such systematic phenomena. It is also concerned to explain the rich and subtle semantic behaviour of common and highly flexible words like the verbs 'do' and 'put' and the prepositions 'at', 'in' and 'to'. Each of these words has uses which are so numerous yet so closely related that they are often described as 'polysemous' rather than ambiguous.

Structural ambiguity occurs when a phrase or sentence has more than one underlying structure, such as the phrases 'Tibetan history teacher', 'a student of high moral principles' and 'short men and women', and the sentences 'The

girl hit the boy with a book' and 'Visiting relatives can be boring'. These ambiguities are said to be structural because each such phrase can be represented in two structurally different ways, for example '[Tibetan history] teacher' and 'Tibetan [history teacher]'. Indeed, the existence of such ambiguities provides strong evidence for a level of underlying syntactic structure (see **Syntax**). Consider the structurally ambiguous sentence, 'The chicken is ready to eat', which could be used to describe either a hungry chicken or a cooked chicken. It is arguable that the operative reading depends on whether or not the implicit subject of the infinitive clause 'to eat' is tied anaphorically to the subject ('the chicken') of the main clause.

It is not always clear when we have a case of structural ambiguity. Consider the elliptical sentence, 'Perot knows a richer man than Trump'. It has two meanings, that Perot knows a man who is richer than Trump and that Perot knows a man who is richer than any man Trump knows, and is therefore ambiguous. But what about the sentence 'John loves his mother and so does Bill'? It can be used to say either that John loves John's mother and Bill loves Bill's mother or that John loves John's mother and Bill loves John's mother. But is it really ambiguous? One might argue that the clause 'so does Bill' is unambiguous and may be read unequivocally as saying in the context that Bill does the same thing that John does, and although there are two different possibilities for what counts as doing the same thing, these alternatives are not fixed semantically. Hence the ambiguity is merely apparent and better described as semantic underdetermination.

Although ambiguity is fundamentally a property of linguistic expressions, people are also said to be ambiguous on occasion in how they use language. This can occur if, even when their words are unambiguous, their words do not make what they mean uniquely determinable. Strictly speaking, however, ambiguity is a semantic phenomenon, involving linguistic meaning rather than speaker meaning (see **Meaning and communication**). Generally when one uses ambiguous words or sentences, one does not consciously entertain their unintended meanings, although there is psycholinguistic evidence that when one hears ambiguous words one momentarily accesses and then rules out their irrelevant senses. When people use ambiguous language, generally its ambiguity is not intended. Occasionally, however, ambiguity is deliberate, as with an utterance of 'I'd like to see more of you' when intended to be taken in more than one way in the very same context of utterance.

2 Ambiguity contrasted

It is a platitude that what your words convey 'depends on what you mean'. This suggests that one can mean different things by what one says, but it says nothing about the variety of ways in which this is possible. Semantic ambiguity is one such way, but there are others: homonymy (mentioned in §1), vagueness, relativity, indexicality, nonliterality, indirection and inexplicitness. All these other phenomena illustrate something distinct from multiplicity of linguistic meaning.

An expression is vague if it admits of borderline cases (see **Vagueness**). Terms like 'bald', 'heavy' and 'old' are obvious examples, and their vagueness is explained by the fact that they apply to items on fuzzy regions of a scale. Terms that express cluster concepts, like 'intelligent', 'athletic' and 'just', are vague because their instances are determined by the application of several criteria, no one of which is decisive.

Relativity is illustrated by the words 'heavy' and 'old' (these are vague as well). Heavy people are lighter than nonheavy elephants, and old cats can be younger than some young people. A different sort of relativity occurs with sentences like 'Jane is finished' and 'John will be late'. Obviously one cannot be finished or late *simpliciter* but only finished with something or late for something. This does not show that the words 'finished' and 'late' are ambiguous (if they were, they would be ambiguous in as many ways as there are things one can be finished with or things one can be late for), but only that such a sentence is semantically underdeterminate - it must be used to mean more than what the sentence means.

Indexical terms, like 'you', 'here' and 'tomorrow', have fixed meaning but variable reference. For example, the meaning of the word 'tomorrow' does not change from one day to the next, though of course its reference does (see **Demonstratives and indexicals**).

Nonliterality, indirection and inexplicitness are further ways in which what a speaker means is not uniquely determined by what their words mean (see **Speech acts §4**). They can give rise to unclarity in communication, as might happen with utterances of 'You're the icing on my cake', 'I wish you could sing longer and louder', and 'Nothing is on television tonight'. These are not cases of linguistic ambiguity but can be confused with it because

speakers are often said to be ambiguous.

3 Philosophical relevance

Philosophical distinctions can be obscured by unnoticed ambiguities. So it is important to identify terms that do double duty. For example, there is a kind of ambiguity, often described as the ‘act/object’ or the ‘process/product’ ambiguity, exhibited by everyday terms like ‘building’, ‘shot’ and ‘writing’. Confusions in philosophy of language and mind can result from overlooking this ambiguity in terms like ‘inference’, ‘statement’ and ‘thought’. Another common philosophical ambiguity is the type/token distinction. Everyday terms like ‘animal’, ‘book’ and ‘car’ apply both to types and to instances (tokens) of those types. The same is true of linguistic terms like ‘sentence’, ‘word’ and ‘letter’ and of philosophically important terms like ‘concept’, ‘event’ and ‘mental state’ (see [Type/token distinction](#)).

Although unnoticed ambiguities can create philosophical problems, ambiguity is philosophically important also because philosophers often make spurious claims of it. Indeed, the linguist Charles Ruhl (1989) has argued that certain ostensible ambiguities, including act/object and type/token, are really cases of lexical underdetermination. [Kripke \(1977\)](#) laments the common stratagem, which he calls ‘the lazy man’s approach in philosophy’, of appealing to ambiguity to escape from a philosophical quandary, and [Grice \(1967\)](#) urges philosophers to hone ‘Modified Occam’s Razor: senses are not to be multiplied beyond necessity’. He illustrates its value by shaving a sense off the logical connective ‘or’, often thought to have both an inclusive and exclusive sense. Grice argues that, given its inclusive meaning, its exclusive use can be explained entirely on pragmatic grounds (see [Implicature §6](#); [Pragmatics §12](#)). Another example, prominent in modern philosophy of language, is the ambiguity alleged to arise from the distinction between referential and attributive uses of definite descriptions (see [Descriptions §5](#)). Less prominent but not uncommon is the suggestion that pronouns are ambiguous as between their anaphoric and deictic use. So, for example, it is suggested that a sentence like ‘Oedipus loves his mother’ has two ‘readings’ - that is, it is ambiguous - because it can be used to mean either that Oedipus loves his own mother or that Oedipus loves the mother of some contextually specified male. However, this seems to be an insufficient basis for the claim of ambiguity. After all, being previously mentioned is just another way of being contextually specified. Accordingly, there is nothing semantically special in this example about the use of ‘his’ to refer to Oedipus.

Claims of structural ambiguity can also be controversial. Of particular importance are claims of scope ambiguity, which are commonly made but rarely defended (see [Scope](#)). A sentence like ‘Everybody loves somebody’ is said to exhibit a scope ambiguity because it can be used to mean either that for each person, there is somebody that that person loves or (however unlikely) that there is somebody that everybody loves. These uses may be represented, respectively, by the logical formulas ‘ $(\forall x)(\exists y)(Lxy)$ ’ and ‘ $(\exists y)(\forall x)(Lxy)$ ’. It is generally assumed that, because different logical formulas are needed to represent the different ways in which an utterance of such a sentence can be taken, the sentence itself has two distinct logical forms (see [Logical form](#)). Sustaining this claim of ambiguity requires identifying a level of linguistic description at which the sentence can be assigned two distinct structures. Some grammarians have posited a level of LF, corresponding to what philosophers call logical form, at which relative scope of quantified noun phrases may be represented. However, LF of this kind does not explain scope ambiguities that philosophers attribute to sentences containing modal operators and psychological verbs, such as ‘The next president might be a woman’ and ‘Ralph wants a sloop’. An utterance of such a sentence can be taken in either of two ways, but it is arguable that the sentence is not ambiguous but merely semantically underdeterminate with respect to its two alleged ‘readings’.

Notwithstanding the frequency in philosophy of unwarranted and often arbitrary claims of ambiguity, it cannot be denied that some terms really are ambiguous. The nouns ‘bank’ and ‘suit’ are clear examples and so are the verbs ‘bank’ and ‘file’. Philosophers sometimes lament the prevalence of ambiguity in natural languages and yearn for an ideal language in which it is absent. But ambiguity is a fact of linguistic life. Despite the potentially endless supply of words, many words do double duty or more. And despite the unlimited number of sentences, many have several meanings and their utterance must be disambiguated in the light of the speaker’s likely intentions.

See also: [Language, philosophy of](#); [Semantics](#)

KENT BACH

References and further reading

- Atlas, J.D.** (1989) *Philosophy Without Ambiguity: A Logico-Linguistic Essay*, Oxford: Oxford University Press. (Examines ambiguity tests and questions certain philosophical appeals to ambiguity.)
- Bach, K.** (1994) 'Conversational Implicature', *Mind and Language* 9: 124-62. (Identifies ways, distinct from those identified by Grice, in which linguistic meaning can underdetermine speaker meaning and, in particular, contrasts semantic underdetermination with ambiguity.)
- Cruse, D.A.** (1986) *Lexical Semantics*, Cambridge: Cambridge University Press. (Chapter 3 discusses linguistic features of ambiguity and examines tests for it.)
- Grice, H.P.** (1967) 'Logic and Conversation', *Studies in the Way of Words*, Cambridge, MA: Harvard University Press, 1989. (Chapter 2 of this collection explains how multiple uses of an expression can often be explained without appealing to ambiguity.)
- Kripke, S.A.** (1977) 'Speaker's Reference and Semantic Reference', *Midwest Studies in Philosophy* 2: 255-76. (A case study, focusing on the distinction between referential and attributive uses of definite descriptions, in how to expose a gratuitous philosophical claim of ambiguity.)
- May, R.** (1985) *Logical Form: Its Structure and Derivation*, Cambridge, MA: MIT Press. (Investigates the hypothesis within generative grammar that there is a level of representation, Logical Form.)
- Ruhl, C.** (1989) *On Monosemy: A Study in Linguistic Semantics*, Albany, NY: State University of New York Press. (Argues for the presumption that a word has a single, though highly abstract, meaning rather than the multiplicity of meanings commonly attributed, as by the list of definitions in its dictionary entry.)
- Zwicky, A. and Sadock, J.** (1975) 'Ambiguity Tests and How to Fail Them', in J. Kimball (ed.) *Syntax and Semantics*, vol. 4, New York: Academic Press. (Presents and assesses various linguistic tests of ambiguity.)

Analysis, nonstandard

Nonstandard analysis is an important application of mathematical logic to the rest of mathematics. Invented in 1960, it provided a long-sought-for rigorous justification for the use of infinitely large and infinitely small (infinitesimal) quantities in the differential and integral calculus, and the first sound canon for manipulating such quantities.

Consider the structure $\mathbb{R}$ of real numbers, that is, the set of real numbers together with operations and relations on them. We specify a formal language $\mathcal{L}$, which has names for all individual real numbers as well as for the operations and relations of $\mathbb{R}$. We can then obtain an enlargement (a special kind of extension) ${}^*\mathbb{R}$ ('pseudo- $\mathbb{R}$ ') of $\mathbb{R}$ such that ${}^*\mathbb{R}$ and $\mathbb{R}$ have exactly the same formal properties: that is, properties expressible in $\mathcal{L}$. But, far from being merely a copy of $\mathbb{R}$, ${}^*\mathbb{R}$ has convenient properties not expressible in $\mathcal{L}$, which are not shared by $\mathbb{R}$. In particular, ${}^*\mathbb{R}$ has additional 'nonstandard' objects (which have no names in $\mathcal{L}$), some of which behave as infinite or infinitesimal quantities. If, using these novel objects and properties of ${}^*\mathbb{R}$, we prove a proposition about ${}^*\mathbb{R}$ that can be expressed in $\mathcal{L}$, then this proposition holds automatically also for $\mathbb{R}$. Such 'nonstandard' proofs of results about $\mathbb{R}$ are often easier and more intuitive than conventional proofs, which operate wholly within $\mathbb{R}$.

More generally, this method is applied to other structures, in virtually every branch of mathematics - including algebraic number theory, various branches of classical and modern analysis, probability theory and mathematical physics, to yield highly intuitive characterizations of various infinitary concepts, to simplify proofs and to provide new, efficient ways of generating mathematical constructs.

1 Historical background

The lineage of nonstandard analysis (NSA) goes back to Leibniz in two ways: first, his advocacy of the use of infinite and infinitesimal quantities in analysis was remarkably sophisticated for his time (see [Analysis, philosophical issues in §1](#)); and second, his vision of *lingua characteristica* and *calculus ratiocinator* foreshadowed the invention of mathematical logic, which was to provide the means for a rigorous justification of the use of those quantities.

Leibniz regarded infinite and infinitesimal quantities as 'useful fictions'; ideal (rather than metaphysically real) entities, whose status was similar to that of imaginary roots such as $\sqrt{-2}$, and whose adjunction to the standard real numbers serves to shorten proofs and aid mathematical invention in problems involving continuity.

Like complex (and, in particular, imaginary) numbers, infinite and infinitesimal quantities were freely used by mathematicians during the eighteenth century, despite known inconsistencies apparently inherent in this practice, which provoked protests by philosophical critics such as Berkeley.

During the nineteenth century, with the development of a more critical attitude to the foundations of analysis, the notions of the real and complex number structures were greatly clarified, and consistent canons were established for operating with these entities (see [Numbers](#)). However, no rigorous canon was found for the use of infinite and infinitesimal quantities. (Cantor, the inventor of set theory, thought that he could actually *prove* that no such canon was possible, and his view was accepted by Peano and others.) Instead, work on the foundations of analysis - by Bolzano, Weierstrass, Dedekind and others - followed a different route (whose viability had also been asserted by Leibniz): the 'delta-epsilon' method, whereby 'instead of the infinite or infinitely small, one takes [standard] quantities as large or as small as required to make the error smaller than [any] given error' (see [Analysis, philosophical issues in §2](#)).

By the twentieth century, the success of this method (in effect, an extension of Archimedes' method of exhaustion) had been so spectacular and its domination in pure mathematics so total, that Russell was merely expressing received opinion when he asserted that 'infinitesimals as explaining continuity must be regarded as unnecessary, erroneous, and self-contradictory' ([1903] 1937: 345). Nevertheless, the older, now officially discredited, method survived in 'lowbrow' (mainly applied) mathematics and - as a powerfully creative heuristic aid - in the 'underground' private deliberations of 'highbrow' mathematicians.

Before the advent of NSA, some isolated attempts were still made, in the face of received opinion, to provide a

rigorous justification for the use of infinite and infinitesimal quantities in analysis. None of these was fully successful, although Schmieden and Laugwitz (1958) came quite close. The successful solution of this problem, achieved by NSA, was to use tools produced by mathematical logic during the first half of the twentieth century, in particular, the notion of a nonstandard model. Skolem showed the existence of nonstandard models of axiomatic set theory (1923) and of arithmetic (1934), but for a long time such models were regarded by logicians as pathological ‘monsters’ rather than a serious object of study - let alone a useful tool. This attitude began to change in about 1949, and during the 1950s there was a growing body of research into the properties and inner make-up of nonstandard models of arithmetic. This set the stage for Abraham Robinson’s invention, in the autumn of 1960, of NSA, which uses nonstandard models as an instrument.

2 Enlargements

The method consists in systematically exploiting the peculiar properties of an enlargement (a special kind of nonstandard extension) ${}^*\mathcal{U}$ (pronounced ‘pseudo- $\mathcal{U}$ ’) of a mathematical structure $\mathcal{U}$, in order to study $\mathcal{U}$ itself.

As a paradigmatic example, consider the structure $\mathcal{R}$ whose domain of individuals is the set $\mathbb{R}$ of real numbers. Apart from the domain $\mathbb{R}$ $\mathcal{R}$ has as constituents all relations and operations on this domain. Let $\mathcal{L}$ be a first-order formal language which can be interpreted in $\mathcal{R}$: so that the domain $\mathbb{R}$ can be taken as the range of the variables of $\mathcal{L}$, and for each entity of $\mathcal{R}$ - an individual real, or a relation among reals, or an operation on reals - $\mathcal{L}$ has a symbol intended as a name for that entity.

Using the ‘compactness theorem’, we can prove the existence of a structure ${}^*\mathcal{R}$, satisfying two conditions. First, ${}^*\mathcal{R}$ is an ‘elementary extension’ of $\mathcal{R}$. Second (‘weak saturation condition’), if Φ is any family of $\mathcal{L}$ -formulas having one free variable x , which is finitely satisfiable in $\mathcal{R}$ (that is, every finite subset of Φ is satisfied by some individual real), then Φ as a whole is satisfiable in ${}^*\mathcal{R}$.

The elementary extension condition is tantamount to the following ‘transfer principle’: any $\mathcal{L}$ -sentence is true (when interpreted) in $\mathcal{R}$ if and only if it is true (when re-interpreted) in ${}^*\mathcal{R}$.

Thus any property of $\mathcal{R}$ that can be formalized by means of an $\mathcal{L}$ -sentence is true also of ${}^*\mathcal{R}$, and conversely. For example, many well-known arithmetical laws (such as the fact that $x + y = y + x$ for any reals x and y) are automatically transferred to the extended domain ${}^*\mathbb{R}$ of the structure ${}^*\mathcal{R}$. Similarly, the fact that the relation denoted by ‘ $<$ ’ is a total (that is, linear) ordering of $\mathbb{R}$ also transfers to ${}^*\mathbb{R}$. In this sense, $\mathcal{R}$ and ${}^*\mathcal{R}$ have exactly the same formal properties.

The members of ${}^*\mathbb{R}$ are called ‘ * reals’ (‘pseudo-reals’). Note that every (standard) real (that is, member of $\mathbb{R}$) is also a * real, but not conversely. A * real which does not belong to $\mathbb{R}$ is said to be ‘nonstandard’.

As an illustration of the (weak) saturation condition, consider the family of all $\mathcal{L}$ -formulas of the form $r < x$, where x is a variable (common to all these formulas) and r is any $\mathcal{L}$ -constant denoting a real r . Clearly, this family of formulas is finitely satisfiable in $\mathbb{R}$ because for any finitely many reals $r_1, r_2, \dots, r_n$ there is always another real greater than all of them. Hence by saturation there exists some - necessarily nonstandard - * real s such that $r < s$ for every real r . Such an s , being greater than all reals, is to be regarded as a ‘positive infinite’ quantity. Similarly, there are nonstandard ‘negative infinite’ * reals. The reciprocals of infinite * reals, as well as the real 0, are ‘infinitesimals’, which are smaller than all positive reals and greater than all negative ones.

If s and t are * reals such that $s - t$ is infinitesimal, we write ‘ $s \approx t$ ’ and say that s and t are ‘infinitely close’ to each other. In whimsical homage to Leibniz, the collection $\mu(r)$ of all * reals infinitely close to a given real r was termed by Robinson the ‘monad’ of r .

By referring to the entities of an enlargement ${}^*\mathcal{R}$, many standard concepts of analysis receive intuitively appealing characterizations which, while being rigorous, are often strikingly similar to the definitions used (without rigorous justification) before the advent of the ‘delta-epsilon’ method. For example, let f be a real-valued function defined in the neighbourhood of a real r . Then f is continuous at r if and only if $f(s) \approx f(r)$ whenever $s \approx r$. Also, f is differentiable at r if and only if there is a unique real d such that

$$\frac{f(r+\delta)-f(r)}{\delta} \approx d$$

for every non-zero infinitesimal δ . This d is then the derivative of f at r .

Such nonstandard characterizations of standard concepts can be used to obtain nonstandard proofs of standard results concerning $\mathbb{R}$ itself. NSA is ‘conservative’, in the sense that any standard result that can be proved using NSA can, at least in principle, be proved also by purely standard means. However, the nonstandard proof is often considerably shorter and more transparent - and therefore easier to construct - than a conventional standard proof. Moreover, as shown by Henson and Keisler (1986), NSA is essentially more powerful than traditional standard analysis. (See also Keisler 1994.)

The ideas outlined so far in connection with the first-order structure $\mathbb{R}$ of reals apply quite generally to any first-order mathematical structure $\mathcal{U}$. Any such $\mathcal{U}$ can be taken as the ‘ground’ structure: we can prove the existence of an enlargement ${}^*\mathcal{U}$ of $\mathcal{U}$ satisfying the same two conditions (elementary extension and weak saturation).

3 Higher-order and special enlargements

Modern analysis is largely concerned with higher-order entities such as arbitrary sets of reals, sets of real-valued functions, sets of sets of functions, sets of sets of reals, functions from sets of reals to sets of reals, and so on.

To apply NSA in this wider context, Robinson used (from 1962) a type-theoretic approach: he considered a ‘layered’ type-theoretic structure which, in addition to the ground-level individuals, say, the reals, also contains all entities of finite orders obtainable from these individuals (such as those mentioned in the preceding paragraph). By a well-known procedure, a layered type-theoretic structure can be replaced by a first-order structure. Essentially, this amounts to considering the higher-order entities as ground-level individuals of different sorts. The membership relation between entities of different orders now becomes a first-order relation connecting individuals of different sorts. This many-sorted first-order structure can be taken as the ground structure $\mathcal{U}$ whose enlargements are obtained as outlined in §2 above.

An equivalent but technically simpler approach, first proposed by Machover (1967), is to embed the original set of individuals in a model of Zermelo’s set theory - a sufficiently large part of the so-called ‘cumulative hierarchy of sets’ (see [Set theory §1](#)). This constitutes a first-order structure in which all the entities of analysis are present as individuals (of the same sort). The basic relations in such a Zermelo structure are the first-order relations of equality ($=$) and set membership ($\in$). Zermelo structures can be enlarged in the usual way.

Another set-theoretic approach, invented by E. Nelson (1977), proceeds axiomatically. Nelson’s internal set theory is a conservative extension of Zermelo-Fraenkel set theory with the axiom of choice (see [Set theory §1](#)). Its universe of discourse contains all standard sets as well as nonstandard * sets. Using a special formal primitive predicate, ‘is standard’, counterparts of the transfer principle and saturation conditions are decreed as additional postulates.

For many applications of NSA it is irrelevant which of the infinitely many enlargements of a given ground structure is used. But special enlargements are sometimes singled out either for convenience of exposition or because their special properties are needed for a particular application.

One special kind of enlargement is obtained by taking an ‘ultrapower’ of the ground structure - an infinitary construction foreshadowed by Skolem in 1934 for the natural numbers and Hewitt (1948) for the reals. Luxemburg used an ultrapower enlargement of $\mathbb{R}$ in one of the earliest broad expositions of NSA (1962). The advantage of this lucid presentation is that it does not presuppose acquaintance with logical technicalities, and was therefore particularly suited for an audience of mathematicians unfamiliar with them. However, it paid for this convenience by having to prove separately each instance of the transfer principle.

Some applications of NSA need enlargements that satisfy certain stronger saturation conditions. Notably, the extremely useful procedure for obtaining a Loeb measure (see below) requires the enlargement to satisfy a condition known as ‘ $\aleph_1$ -saturation’.

4 Range of applications

Robinson himself initiated a formidable array of applications of NSA in a wide spectrum of mathematical topics,

ranging from algebraic number theory, through various branches of classical and modern analysis and probability theory, to mathematical physics. Here we can mention only a few applications.

Generally speaking, NSA is a particularly powerful tool in problems involving the topological notion of compactness, because the nonstandard characterization of compactness is extremely simple and transparent compared with the standard definition.

Indeed, one of the earliest impressive successes of NSA was the solution by Robinson (1966) with A.R. Bernstein of an open problem concerning Hilbert spaces, which involved topological compactness in an essential way. Since then there have been many important applications to linear spaces and functional analysis.

NSA also provides illuminating insights into structures in which continuity or analyticity and algebraic notions interact; for example, in topological and analytic groups, whose nonstandard study was initiated by Robinson. In such structures, the whole of the continuous or analytic aspect is ‘algebraicized’ and encapsulated in a single group of infinitesimal ^{*} members. Exploiting these ideas, Hirschfeld (1990) was able to simplify very significantly the solution of one of Hilbert’s celebrated problems.

Some of the most spectacular achievements of NSA are in probability theory. In this connection, E. Nelson’s work (1987) is noteworthy in providing a far-reaching simplification of difficult parts of the theory. Arguably the most important breakthrough in NSA published after Robinson’s death was the invention by Loeb (1975) of a powerful nonstandard procedure for obtaining standard probability measures (as well as more general unbounded measures) with prescribed desirable properties.

Robinson himself was - unusually for a logician - an expert in applied mathematics and mathematical physics (particularly in fluid dynamics). Applications of NSA in these areas, initiated by him and continued by others, have been of very great importance.

5 Status of NSA

If revolutions have ever occurred in mathematics, then NSA may be described as one of them (Dauben 1992). Logicians have in general welcomed it, in some notable cases with great enthusiasm. Thus, in Gödel’s view, ‘there are good reasons to believe that nonstandard analysis, in some version or other, will be the analysis of the future’ (see Robinson [1966] 1974). Some logicians have written elementary calculus text-books for beginners using NSA (for example, Keisler 1976).

Among working mathematicians generally, the spread of NSA techniques has been relatively slow. As pointed out in §2, any nonstandard proof of a standard theorem can theoretically be replaced by a standard proof. Although nonstandard techniques are often significantly more efficient, they require familiarity with logic, which many mathematicians do not possess. However, the acceptance of NSA among mathematicians has been boosted by the successes of its practitioners. In this respect the results initiated by Loeb’s work (see §4) have played a pivotal role.

But the view that NSA is destined to supersede standard analysis altogether should be taken with a grain of salt. This is because a given structure has infinitely many enlargements, none of which can be singled out as in any way canonical or ‘best’, and a standard definition is needed in order to verify that a nonstandard characterization of a concept is independent of the choice of enlargement (see Machover 1993).

See also: [Logical and mathematical terms, glossary of](#)

MOSHÉ MACHOVER

References and further reading

- Capiński, M. and Cutland, N.** (1992) ‘Stochastic Navier-Stokes Equations’, *Acta Applicanda Mathematicae* 25: 59-85. (Powerful use of Loeb measure and other NSA tools to solve a long-standing existence problem in applied mathematics.)
- Cutland, N.** (ed.) (1988) *Nonstandard Analysis and its Applications*, Cambridge: Cambridge University Press. (Collection of papers on NSA, including a wide range of applications to pure and applied mathematics. See especially S. Albeverio, ‘Applications of Nonstandard Analysis in Mathematical Physics’, and C.W. Henson, ‘Infinitesimals in Functional Analysis’.)

- Dauben, J.W.** (1992) 'Revolutions Revisited', appendix to 'Conceptual Revolutions and the History of Mathematics: Two Studies in the Growth of Knowledge', in D. Gillies (ed.) *Revolutions in Mathematics*, Oxford: Clarendon Press. (Assessment of NSA as 'a contemporary revolution in mathematics'; mentioned in §5.)
- Dauben, J.W.** (1994) *Abraham Robinson: The Creation of Nonstandard Analysis, a Personal and Mathematical Odyssey*, Princeton, NJ: Princeton University Press. (An authoritative scientific biography.)
- Henson, C.W. and Keisler, H.J.** (1986) 'The Strength of Nonstandard Analysis', *Journal of Symbolic Logic* 51: 377-86. (Explores the strength of NSA compared to conventional analysis; mentioned in §2.)
- Hewitt, E.** (1948) 'Rings of Real-Valued Continuous Functions I', *Transactions of the American Mathematical Society* 64: 54-99. (Construction of a structure that, in retrospect, is seen to be essentially an ultrapower enlargement of $\mathbb{R}$; mentioned in §3.)
- Hirschfeld, J.** (1990) 'The Nonstandard Treatment of Hilbert's Fifth Problem', *Transactions of the American Mathematical Society* 321: 379-400. (Great simplification compared to the original 1955 solution by D. Montgomery and L. Zippin; mentioned in §4.)
- Keisler, H.J.** (1976) *Elementary Calculus*, Boston, MA: Prindle, Weber & Schmidt. (A beginner's calculus course using NSA; mentioned in §5.)
- Keisler, H.J.** (1994) 'The Hyperreal Line', in P. Ehrlich (ed.) *Real Numbers, Generalizations of the Reals, and Theories of Continua*, Dordrecht: Kluwer. (A masterly survey of NSA as applied to the real line. Discusses the philosophical status of the reals. Mentioned in §2.)
- Lindström, T.** (1988) 'An Invitation to Nonstandard Analysis', in N. Cutland (ed.) *Nonstandard Analysis and its Applications*, Cambridge: Cambridge University Press. (A very user-friendly introduction to NSA. Extensive bibliography up to the late 1980s.)
- Loeb, P.A.** (1975) 'Conversion from Nonstandard to Standard Measure and Applications in Probability Theory', *Transactions of the American Mathematical Society* 211: 113-22. (Construction of the Loeb measure. Pivotal advance in NSA. Mentioned in §4.)
- Luxemburg, W.A.J.** (1962) *Non-Standard Analysis: Lectures on A. Robinson's Theory of Infinitesimals and Infinitely Large Numbers*, Pasadena, CA: California Institute of Technology. (Early extensive presentation of NSA; uses ultrapower construction of enlargement but no formal logic; mentioned in §3.)
- Machover, M.** (1967) *Nonstandard Analysis Without Tears: An Easy Introduction to A. Robinson's Theory of Infinitesimals*, Technical Report no. 27, US Office of Naval Research, Jerusalem: The Hebrew University. (Introduces a set-theoretic framework for applying NSA to higher-order entities; mentioned in §3.)
- Machover, M.** (1993) 'The Place of Nonstandard Analysis in Mathematics and in Mathematics Teaching', *British Journal for the Philosophy of Science* 44: 205-12. (Expands on §5.)
- Nelson, E.** (1977) 'Internal Set Theory: A New Approach to Nonstandard Analysis', *Bulletin of the American Mathematical Society* 83: 1,165-98. (Elegant axiomatic framework for NSA; mentioned in §3.)
- Nelson, E.** (1987) *Radically Elementary Probability Theory*, Annals of Mathematics Studies 117, Princeton, NJ: Princeton University Press. (Tour de force of simplification: advanced probability theory brilliantly made easy by NSA; mentioned in §4.)
- Robinson, A.** (1961) 'Non-Standard Analysis', *Proceedings of the Royal Academy of Sciences, Amsterdam*, series A, 64: 432-40. (The first publication on NSA.)
- Robinson, A.** (1962) *Complex Function Theory over Non-Archimedean Fields*, Technical-Scientific Note no. 30, US Air Force Office of Scientific Research, Jerusalem: The Hebrew University. (NSA applied for the first time extensively to higher-order entities, using a type-theoretic framework, to prove new results on the zeros of complex polynomials and on the behaviour of an analytic function near an essential singularity; mentioned in §3.)
- Robinson, A.** (1966) *Non-Standard Analysis*, Amsterdam: North Holland; 2nd revised edn, 1974. (Uses a type-theoretic framework for higher-order entities; includes a wealth of applications; and presents the result on Hilbert spaces, mentioned in §4, proved with A.R. Bernstein. A historical chapter discusses in some detail topics touched upon in §1 and quotes extensively from Leibniz and others. Gödel's preface to the 1974 edition gives his assessment of NSA quoted in §5.)
- Robinson, A.** (1979) *Selected Papers of Abraham Robinson*, vol. 2, *Nonstandard Analysis and Philosophy*, ed. H.J. Keisler, W.A.J. Luxemburg and S. Körner, New Haven, CT: Yale University Press. (Includes Robinson's papers on NSA, including seminal applications in diverse fields.)
- Russell, B.A.W.** (1903) 'Philosophical Arguments concerning the Infinitesimal', in *The Principles of Mathematics*, London: Cambridge University Press, 2nd edn, 1937; repr. London: Routledge, 1992. (Quoted in §1.)

- Schmieden, C. and Laugwitz, D.** (1958) 'Eine Erweiterung der Infinitesimalrechnung' (An Extension of the Infinitesimal Calculus), *Mathematische Zeitschrift* 69: 1-39. (The last, partly successful, pre-NSA attempt to justify infinitesimals.)
- Skolem, T.** (1923) 'Einige Bemerkungen zur axiomatischen Begründung der Mengenlehre', in *Matematikerkongressen i Helsingfors den 4-7 Juli 1922, Den femte skandinaviska matematikerkongressen, Redogörelse*, Helsinki: Akademiska Bokhandeln, 217-32; trans. S. Bauer-Mengelberg, 'Some Remarks on Axiomatized Set Theory', in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967, 291-301. (Proves the existence of nonstandard models of axiomatic set theory and hints at the existence of nonstandard models of arithmetic; mentioned in §1.)
- Skolem, T.** (1934) 'Über die nicht-charakterisierbarkeit der Zahlenreihe mittels endlich oder abzählbar unendlich vieler Aussagen mit ausschliesslich Zahlenvariablen' (On the Non-Characterizability of the Number Sequence by Finitely or Denumerably Many Statements with Number Variables Only), *Fundamenta Mathematicae* 23: 150-61. (Existence of nonstandard models of arithmetic proved using a construction foreshadowing ultrapowers.)

Analysis, philosophical issues in

The term ‘mathematical analysis’ refers to the major branch of mathematics which is concerned with the theory of functions and includes the differential and integral calculus. Analysis and the calculus began as the study of curves, calculus being concerned with tangents to and areas under curves. The focus was shifted to functions following the insight, due to Leibniz and Isaac Newton in the second half of the seventeenth century, that a curve is the graph of a function. Algebraic foundations were proposed by Lagrange in the late eighteenth century; assuming that any function always took an expansion in a power series, he defined the derivatives from the coefficients of the terms. In the 1820s his assumption was refuted by Cauchy, who had already launched a fourth approach, like Newton’s based on limits, but formulated much more carefully. It was refined further by Weierstrass, by means which helped to create set theory. Analysis also encompasses the theory of limits and of the convergence and divergence of infinite series; modern versions also use point set topology. It has taken various forms over the centuries, of which the older ones are still represented in some notations and terms. Philosophical issues include the status of infinitesimals, the place of logic in the articulation of proofs, types of definition, and the (non-) relationship to analytic proof methods.

1 The calculus: Newton, Leibniz and Lagrange

To begin in geometric terms, the differential calculus provides methods of determining the slope of a tangent to a curve, while the integral calculus finds the area between the curve and the x -axis. The study of tangents and areas goes back to ancient times, but prior to Isaac Newton and G.W. Leibniz tangents were determined at specific points and areas were always definite ones, that is, taken between a specific starting point and end point on the x -axis. Their new insight - treating a curve as being the graph of a function $f(x)$ - was to determine the tangents and areas as new *functions*. Thus the derivative function f' gives the rate of change of the function f , and the indefinite integral $\int f$ is a function showing how the area under f (taken from some arbitrary starting point) varies with the end point. The two calculi are inverse: the ‘fundamental theorem’ states that

$$(1) (\int f)' = f \text{ and also } \int(f') = f (+ \text{constant}).$$

In Newton’s version (created 1660s, published 1700s; see Newton (1967-81); [Newton, I. §6](#)) the differential was a ‘fluxion’, the rate of change of a variable quantity relative to time (understood conventionally); conversely, that variable was the ‘fluent’ of its fluxion. His definitions involved changes in dimension: a point as the limit of a sequence of lines of decreasing length, and so on. He defined his functions in terms of limits, but in a very naïve way.

Leibniz formed his differential and integral calculus in the 1670s (published 1680s; see [Leibniz, G.W. §§1, 10](#)) independently of Newton, and in a quite different form. It was based on the idea that a ‘differential’ dx could be created as an infinitesimal increment on x , with the same dimension as x ; the slope of the tangent to the curve corresponding to $f(x)$ was $df(x)/dx$. For linear variables x and $f(x)$ the ‘integral’ $\int f(x)dx$ was the sum of the areas of rectangles $f(x)$ high and dx wide. Thus his version of the principal relationship (1) is

$$(2) d(\int f(x)dx)/dx = \int(df(x)/dx)dx = f(x).$$

Another point of superiority of these theories over their predecessors was that the defined functions were themselves variable; thus they admitted of higher-order fluxions (differentials) and fluents (integrals). Physical phenomena of certain kinds could be expressed as ‘differential’ (or ‘fluxional’) equations of some order; finding their solution(s) became a major concern in mathematics, with many important applications in mechanics and geometry.

Of the two forms, Leibniz’s gradually gained a considerable ascendancy during the eighteenth century. In one respect his use of the word ‘function’ was misleading, in that often no functional dependence obtained: variables could vary together. But during the 1750s Leonhard Euler realized the advantage of systematically specifying one variable (x , say) as a basis in a problem, relative to which the others were functionally dependent. His criterion was that dx be constant, so that further derivatives $ddx, \dots$ were all zero; then the differentials of other variables (y , say) were related to dx via ‘differential coefficients’ $p(x), q(x), \dots$ of $f(x)$ with respect to x :

$$(3) \quad df(x) = p(x)dx, \\ dd f(x) = q(x)(dx)^2, \dots$$

Another major extension due then to Euler and others was to allow several variables in a problem to be independent; this move greatly increased the range of phenomena which could be represented, in ‘partial differential equations’.

During the late eighteenth century J.L. Lagrange proposed, as an alternative to limits and infinitesimals, that the differential calculus was obtainable from the so-called ‘Taylor series’, the expansion of a function f about any value of x with $h > 0$:

$$(4) \quad f(x+h) = f(x) + hf'(x) + h^2 f''(x)/2! + \dots$$

(This important result had been named after Brook Taylor, although it had already been found by his master Newton and by Leibniz.) The ‘derived functions’ $f'(x)$, $f''(x)$, ... (to use Lagrange’s name and notation for them) were *defined* as the coefficients of powers of h in the expansion (4); in his opinion they could be determined solely from (4) and algebraic manipulations. The integral calculus was based on reversing the operations; from f'' to f' , and so on.

Lagrange’s approach became most widely circulated through the textbook *Théorie des fonctions analytiques* (1797), which was based upon his teaching at the recently founded École Polytechnique in Paris. He also extended his approach and associated with it an allied subject, the calculus of variations, which produced from a given function neighbouring functions which varied slightly from it. This theory, which became of major importance in applications, is also of philosophical interest, in that it requires possible worlds in which, for example, a particle may move along curves defined by functions varying slightly from the function whose curve it actually followed and which the theory was intended to determine.

Lagrange’s approach also came to have profitable influences on algebra in the early nineteenth century; in particular, some of its consequences provided powerful analogies for Boole and De Morgan in their contributions to algebraic logic. But the assumptions behind (4) seemed too good to be true; and in fact (4) was refuted by means of counter-examples, in the 1820s.

2 Mathematical analysis: Cauchy and Weierstrass

These torpedoes came from A.L. Cauchy, who had already begun to develop a fourth approach - ‘mathematical analysis’ (his name) - as part of his own teaching at the École Polytechnique. As with Newton’s calculus, the theory of limits served as his basis, but he treated them in a far more careful way, with general definitions and an exploration of properties. In addition, the theory of functions and the convergence of infinite series joined the calculus under limit theory. He raised the level of rigour in various ways, of which some involved points of logic. He gave independent definitions of the derivative and the integral:

$$f'(x) := \lim[(f(x+h) - f(x))/h] \text{ as } h \rightarrow 0; \\ \int f(x)dx := \lim \sum_r [(x_r - x_{r-1})f(x_{r-1})], \text{ with} \\ x_0 \leq x_{r-1} < x_r \text{ and } x_0 \leq x \leq \lim_{r \rightarrow \infty} x_r, \\ \text{as } \max_r |x_r - x_{r-1}| \rightarrow 0.$$

Cauchy retained Lagrange’s notation and name for $f'(x)$, but otherwise things were quite different. In particular, for the first time, (1) became the fundamental *theorem* of the calculus, concerning the inverse relationship between differentiation and integration under the sufficient conditions that $f(x)$ was continuous (in his sense); previously the switch between the two calculi had been more or less automatic. He showed the merit of systematically specifying necessary and/or sufficient conditions under which theorems were held to be true, and of enhancing theory by seeking weaker versions of the former and stronger versions of the latter. He even pioneered the methodical numbering of expressions in a paper or book.

Although Cauchy’s theory is not very intuitive (his students and colleagues hated it), it gradually became accepted,

especially after its adoption and refinement by Karl Weierstrass (at his widely influential lecture courses at Berlin from the 1860s) as the now standard ‘epsilon-delta’ treatment of limits. He stressed the need for extra care when several variables were moving towards their respective limits simultaneously, for example, in distinguishing between the following:

$$\lim_{x \rightarrow a, r \rightarrow \infty} \sum_r u_r(x), (5) \lim_{x \rightarrow a} \lim_{r \rightarrow \infty} \sum_r u_r(x), \lim_{r \rightarrow \infty} \lim_{x \rightarrow a} \sum_r u_r(x),$$

where in the first case the two limits move together. Weierstrass and his followers also realized the importance of distinguishing the least upper bound of a collection of values from its upper limit. For example, the upper limit of the sequence $5, 7, -4, \frac{3}{2}, \frac{5}{3}, \frac{7}{4}, \dots, (2n-1)/n, \dots$ is 2, but the least upper bound is 7.

Among related developments, Georg Cantor was inspired from the early 1870s to create both his point set theory and transfinite arithmetic from solving certain technical questions in analysis by Weierstrassian methods. The further refinement of conditions on theorems such as (1) included the need to define irrational numbers in terms of rationals in order to prove theorems on their existence without begging the question posed. Cantor and Weierstrass had versions, but the best remembered is the ‘cut’ definition of Richard Dedekind in 1872. Associate each rational number with a point on the continuous straight line by natural order. Cut the line at any point, then the numbers are divided into two classes: to left (L) and to right (R). If L has no greatest member and R no smallest, then the cut does not correspond to any rational number, and thereby defines an irrational one (see [Numbers §6](#)).

The generally rising levels of rigour in proof - especially the increasing numbers of variables in problems, with ever more intricate functional relationships between them - inspired Giuseppe Peano in the late 1880s to create elements of ‘mathematical logic’ (his phrase), in order that the language itself of mathematical analysis would be made more rigorous, by relying less on prose. He formulated the propositional calculus in quite an axiomatic manner (not a common way of doing mathematics at that time), and introduced ‘propositional functions’ and quantification over their argument places. The latter technique was especially significant in clarifying the order of quantifiers for several variables; distinguishing the expressions in (5) is an important example.

The Weierstrassian tradition, together with Cantorian set theory, became standard fare in pure mathematics and some (but by no means all) applications by the beginning of the twentieth century. Cauchy’s theory of the integral was extended into the measure theory of Henri Lebesgue and others in the 1900s. At the same time various forms of the axiom of choice were found, revealing unexpected gaps and assumptions in proof methods; but gradually their place in analysis (and set theory) was evaluated and (for most figures) their non-constructive character was tolerated. Various generalizations were made: for example, defining the calculus over groups of some kind rather than over real numbers. In addition, the mathematical analysis of complex variables, which Cauchy had initiated along with his creation of the real-variable theory, rose to major status in mathematics - in his version and in alternatives due to Bernhard Riemann and to Weierstrass.

In one respect, reaction to Cauchy was negative. He had declared that divergent series were normally illegitimate for mathematical use; however, in the late nineteenth century, it became understood that the sum of a series could only be defined relative to a manner of summing its terms, and that the normal method (first term + second term + ...) was not the only one. Thus a sum such as

$$(6) 1 - 1 + 1 - 1 + \dots = \frac{1}{2}$$

was true under certain methods of summation, although the series had no sum under the usual method of adding term by term. This discovery rendered still more complicated the relationship between series and functions, which had already suffered a jolt from Cauchy’s refutation of Lagrange’s assumption (4).

3 Some philosophical issues

Several points of philosophical interest have already arisen. A few others will be made here.

Various types of infinitesimal had been used in the calculus and analysis over the centuries. The three main kinds, not always clearly distinguished, were: (a) the ‘infinitely small’, a variable quantity with limiting value zero (Cauchy proposed a new form of this kind); (b) ‘nulls’, such as the dx in (2), which are smaller than normal quantities; and (c) ‘pseudo-infinitesimals’ - quantities different from zero and yet smaller than any assignable

quantity (Berkeley's criticism of the foundations of the calculus in 1734 were based upon this view). They all became unacceptable in Weierstrass' reign; but even then they gained some adherents, and since the 1960s theories have been produced which vindicate their use. The best known is the 'nonstandard analysis' of Abraham Robinson (see [Analysis, nonstandard](#)).

One aspect which splits limit theory apart from infinitesimals is that of dimensions. For Leibniz, if x is a line, dx is a very short line, ddx is a very, very short line, and so on; thus dimensions are preserved under the operations of differentiation and integration. By contrast, a theory which uses limits changes dimensions; for example, the limit of a sequence of lines is a point. Intellectually this difference is vast, with profound consequences for applications; it is one source of the non-intuitiveness of mathematical analysis, and applied mathematicians have often preferred to work with differentials, which allow the geometry of a given problem, controlled by dimensions, to be represented and manipulated with incomparable clarity and flexibility.

Once Euler had introduced the differential coefficient (alongside infinitesimals) with (3) above, the notion of function was at centre stage. A major philosophical consideration was that of *generality* of the notion of function. Euler himself had said that variables x and y may be related 'in any way' by a function, but he had in mind only the kinds of functions normally used in his day. The same remark applies to successors in the nineteenth century, such as J.P.G. Dirichlet, though with a somewhat wider realm in mind. Only in the era of measure theory do we find the set-theoretic version which is now standard; that x and y are functionally related without assuming any details about the means. A striking example is the 'characteristic function of the rationals', a type of totally discontinuous function which takes one constant value when x is rational and a different one when x is irrational. Dirichlet had held it to be pathological, as it had no integral in any sense that he could imagine; by contrast, it played a central role in Lebesgue's formulation of measure theory.

It followed from Cauchy's theory of functions that a continuous function may be without a derivative at certain points. Riemann extended this aspect by producing an exotic range of functions with corners and/or discontinuities; and Weierstrass capped it in 1872, with an example of a continuous function which never had a derivative. Results of this kind raised new questions in the traditional discourse of continuity and discontinuity (see [Intuitionism](#)).

The axiomatic treatment of analysis is often based on the completeness axiom - every non-empty bounded set of real numbers has a least upper bound (equivalent to Dedekind's cut principle mentioned in §2 above). The completeness axiom is impredicative and therefore violates Russell's vicious circle principle (see [Constructivism in mathematics](#)).

Cauchy's rigour was based in part upon broad definitions of basic concepts; definability and modes of definitions remained an important topic. In particular, Peano treated them as a central part of his logico-analysis: rules for forming them correctly, and the role in theories of definition under hypothesis. In addition, Cauchy had insisted that a mathematical function always be single-valued (so that, for example, $\sqrt{x}$ would have to be split into $+\sqrt{x}$ and $-\sqrt{x}$); in expressing this feature Peano came in 1897 to formulate the criteria for the existence of a referent to a definite description in mathematics that Bertrand Russell was to propose in 1905 for natural language in his famous paper 'On Denoting' (see [Descriptions](#); [Russell, B.A.W. §9](#)).

The long-suffering word 'analysis' was forced to fulfil still more duties by this subject. The traditional link with algebra was already becoming tenuous by the late eighteenth century, and was weakened further by Cauchy's refutation of (4) as a foundation. In addition, proofs in mathematical analysis from Cauchy onwards are usually synthetic, in that they start out from definitions of basic notions (continuity of a function, say) and/or proven results and finish with the theorem required.

Weierstrass introduced a methodology, with particular bearing upon mathematical analysis: examine all the exceptional and special cases, and do not be content with theorems about 'in general...'. Cantor was led to his set theory by obeying such a stricture (see [Cantor, G. §1](#)).

Do foundations have foundations of their own? Mathematical analysis shows beautiful examples to illustrate this question: Weierstrass underpinned the foundations which Cauchy had provided by refining single-limit theory into multiple-limit theory and the other aspects mentioned above, which Cauchy and his contemporaries had handled freely; but he and his generation took for granted making an infinite selection of members from a set, which was

recognized in the axiom of choice to be an additional question.

See also: [Axiom of choice](#); [Continuum hypothesis](#); [Logical and mathematical terms, glossary of](#)

I. GRATTAN-GUINNESS

References and further reading

- Bos, H.J.M.** (1974) 'Differentials, Higher-Order Differentials and the Derivative in the Leibnizian Calculus', *Archive for History of Exact Sciences* 14: 1-90.(On Leibniz and Euler.)
- Bottazzini, U.** (1986) *The Higher Calculus*, Heidelberg: Springer.(Covers both real and complex analysis, 1750-1900.)
- Cohen, H.** (1883) *Das Princip der Infinitesimal-methode und seine Geschichte (The Principle of the Infinitesimal Method and its History)*, Berlin: Dummer.(A fine historico-philosophical survey.)
- Grattan-Guinness, I.** (ed.) (1980) *From the Calculus to Set Theory, 1630-1910*, London: Duckworth. (Covers only real-variable analysis.)
- Hardy, G.H.** (1949) *Divergent Series*, Oxford: Clarendon Press.(Fine coverage, with historical notes.)
- Hawkins, T.W.** (1970) *Lebesgue's Theory of Integration: Its Origins and Development*, Madison, WI: University of Wisconsin Press; repr. New York: Chelsea, 1975.(On measure theory.)
- Landau, L.** (1930) *Grundlagen der Analysis (Foundations of Analysis)*, Leipzig: Teubner; repr. and trans. New York: Chelsea, 1951.(Classic introductory text.)
- Newton, I.** (1967-81) *The Mathematical Papers of Isaac Newton*, ed. D.T. Whiteside, Cambridge: Cambridge University Press, 8 vols.(Extensive selection of manuscripts and editorial commentary.)
- Yushkevich, A.P.** (1976) 'The Concept of Function up to the Middle of the 19th Century', *Archive for History of Exact Sciences* 16: 37-85.(On the development of functions and real variables.)

Analyticity

In *Critique of Pure Reason* Kant introduced the term ‘analytic’ for judgments whose truth is guaranteed by a certain relation of ‘containment’ between the constituent concepts, and ‘synthetic’ for judgments which are not like this. Closely related terms were found in earlier writings of Locke, Hume and Leibniz. In Kant’s definition, an analytic judgment is one in which ‘the predicate B belongs to the subject A, as something which is (covertly) contained in this concept A’ ([1781/1787] 1965: 48). Kant called such judgments ‘explicative’, contrasting them with synthetic judgments which are ‘ampliative’. A paradigmatic analyticity would be: bachelors are unmarried. Kant assumed that knowledge of analytic necessities has a uniquely transparent sort of explanation. In the succeeding two centuries the terms ‘analytic’ and ‘synthetic’ have been used in a variety of closely related but not strictly equivalent ways. In the early 1950s Morton White (1950) and W.V. Quine (1951) argued that the terms were fundamentally unclear and should be eschewed. Although a number of prominent philosophers have rejected their arguments, there prevails a scepticism about ‘analytic’ and the idea that there is an associated category of necessary truths having privileged epistemic status.

1 The attack on ‘analytic’

‘Analytic’ has been used in a wide variety of ways: truth by conceptual containment and truth whose denial is contradictory (Kant 1781/1787); logical truth (Bolzano 1837; Feigl 1949); truth by definition and logical derivation (Frege 1884; Pap 1958); truth in virtue of form (Schlick 1930-1); truth by definition and logical truth (Carnap 1937, 1947); truth by definition (Ayer 1936); truth based on meaning (Ayer 1936; C.I. Lewis 1944); truth by semantical rule (Carnap 1947); truth in all possible worlds (C.I. Lewis 1944; D.K. Lewis 1969); convertibility into logical truth by substitution of synonyms (Quine 1951); truth by implicit convention (Putnam 1962); and so on. Although related, not all of these uses are equivalent. For example, logical truths are not true by definition (in the sense of explicit definition), but they are trivially true by definition *plus* logic. Furthermore, Gödel’s incompleteness result shows that logical derivability and logical truth are not equivalent (see [Gödel’s theorems §3](#)). Likewise various principles (for example, supervenience principles; see [Supervenience](#)) which are true in all possible worlds seem not to be true by definition plus logic (if ‘definition’ does not include ‘implicit definitions’; see [Definition](#)). Similarly, it may be doubted that correct definitions provide exact synonyms (see §4). Little care has been taken to distinguish these disparate uses and needless confusions have resulted. But, as Strawson and Grice (1956) note, observations of this sort ‘would scarcely amount to a rejection of the distinction [as Quine urges]. They would, rather, be a prelude to clarification’.

In ‘[Two Dogmas of Empiricism](#)’ (1951), Quine went far beyond the call for clarification; he argued that there simply was no distinction between analytic and synthetic truths. Quine’s argument is an enthymeme with roughly the following form: there is no non-circular, purely *empiricist* clarification of the distinction, and therefore there is no distinction at all: ‘That there is such a distinction to be drawn at all is an unempirical dogma of empiricists, a metaphysical article of faith’ (see [Quine, W.V. §8](#)).

2 Extending the attack to ‘necessary’ and ‘a priori’

Quine’s attack did not stop with ‘analytic’. Following the lead of logical positivists, he (wrongly) held that, if there were any such distinctions, the analytic/synthetic distinction would be the same as the necessary/contingent distinction, which in turn would be the same as the a priori/a posteriori distinction. So, on his view, since there is no analytic/synthetic distinction, there is no necessary/contingent distinction and no a priori/a posteriori distinction either. It is an error, however, to equate these distinctions, so these further conclusions do not follow.

Quine’s attack on the necessary/contingent distinction has not convinced many. Most philosophers accept the distinction and invoke it in their work; modal logic and modal metaphysics are thriving subjects (see [Modal logic; Modal logic, philosophical issues in](#)). Do modal notions have non-circular definitions? Although definitions have been suggested, such as Church (1951) and Bealer (1982), there would be nothing unreasonable in holding them to be primitive (Prior and Fine 1977). After all, everyone must take *some* notions to be primitive.

What *epistemic justification* do we have for the necessary/contingent distinction? No doubt Quine was right that it cannot be justified on purely empirical grounds (that is, using only phenomenal experience and/or observation).

But justification is not always purely empirical: logicians, mathematicians, linguists and philosophers rely heavily on *intuitions* in the justification of their theories. (Quine himself relies on intuition in defending his formulations of set theory over others.) And there are no convincing arguments that such use of intuitions is illegitimate. Once one acknowledges intuitive evidence, the necessary/contingent distinction has a straightforward justification: we have a very wide range of robust modal intuitions (for example, the intuition that it is contingent that the number of planets is greater than seven - there could have been fewer), and when such intuitions are taken as evidence, the best theory is one which accepts the distinction at face value.

The necessary/contingent distinction is a metaphysical distinction. The a priori/a posteriori distinction, by contrast, is epistemological. Indeed, Kripke and Putnam have convincingly argued that there are necessary truths (Hesperus is Phosphorus, water is H₂O, and so on) which are impossible to justify a priori (see Kripke, S.A. §3; Putnam, H. §3). While the former distinction might well lack a non-circular analysis, the latter certainly has one: *p* is a piece of a posteriori evidence if and only if *p* is the content of an experience (a phenomenal experience, observation, memory or testimony); *p* is a piece of a priori evidence if and only if *p* is a piece of evidence which is not the content of an experience in the relevant sense. (On the assumption that one's intuitions are evidence, the contents of those intuitions would constitute one's a priori evidence.) A theory has an a posteriori justification if and only if the evidence on which that justification is based is a posteriori; a theory has an a priori justification if and only if the evidence on which that justification is based is a priori. Note that this sort of analysis does not presuppose that any theory has an a priori justification; one could deny that there is any a priori evidence. The point is that, contrary to Quine's allegation, the *notion* of a priori justification has a straightforward analysis (see A posteriori; A priori).

Quine seems to assume that a priori judgments would need to be infallible and unrevisable. But this is mistaken, as is evident from our ongoing a priori theorizing about the logical paradoxes. One of the main traditional lines of thought on the a priori - from Plato to Gödel - recognizes that a priori justification is fallible and holistic, relying on dialectic and/or a priori theory construction.

3 Defending 'analytic' against Quine

Unlike 'necessary', 'analytic' is a technical term. Accordingly, it is legitimate to demand, as Quine (1951) does, that its use be explained. This could be done with examples, but none of the (nonequivalent) notions listed above fits the standard examples perfectly, nor does any stand out as *the* most salient. This leaves the alternative of giving a definition. Quine's view is that *not one* of the historically prominent uses of 'analytic' (except the 'logical truth' use) has a satisfactory definition. But this radical scepticism is unwarranted.

Consider the following definition: a necessarily true sentence is analytic if and only if it may be converted into a logically true sentence by replacing its syntactically simple predicates with predicates which mean exactly the same thing. This seems to respect the idea that analytic truths are a priori and moreover are justifiable in a specially simple way. Quine (1960) would object to this definition by appeal to his thesis of the indeterminacy of translation - the thesis that 'there is no fact of the matter' concerning claims about identity of meaning (see Radical translation and radical interpretation §§1-3). His arguments for this thesis have not convinced many philosophers, however, for they depend on quasi-verificationist or behaviourist premises. Most philosophers reject Quine's scepticism about meaning, realism having become the dominant view. One reason for this is the advent of the broadly Gricean picture (Grice 1989) according to which meaning is analysable in terms of the propositional attitudes. Accordingly, if there is a fact of the matter about the latter, there is about the former (see Communication and intention). And, since the cognitivist revolution in psychology and philosophy of mind, nearly everyone is a realist about propositional attitudes. Thus, at least one use of 'analytic sentence' ought to be acceptable to these realist philosophers.

The same moral holds for at least certain uses of 'analytic' as it applies to propositions. Despite Quine's scepticism, most philosophers have become convinced that in logic, psychology and semantics there is need for structured propositions, that is, propositions which have a logical form (or sense structure). This makes possible a definition of 'analytic' in another one of its standard uses (Katz 1986; Bealer 1982): *p* is analytic iff_{def} every proposition having the same form (structure) as *p* is necessary.

4 More damaging problems with ‘analytic’

Although the above definition of ‘analytic’ is cogent, the term so-defined fails to apply to a number of examples which traditionally would have been deemed ‘analytic’; the definition is too narrow. For example, the defined use does not cover Kant’s paradigm example of an analyticity, namely, that bodies are extended. To accommodate this and a wide array of other examples (that circles are curves and so forth; see below), one must turn to a wider definition of ‘analytic’ - one relying on some philosophically robust notion such as definition, conceptual analysis, or the kind of meaning relations which hold between a definiendum and a definiens or between an analysandum and an analysans. For example, it is at least credible that there is a definition of ‘body’ in one of its senses according to which ‘Bodies are extended’ would be true by definition. Unfortunately, these wider accounts of ‘analytic’ give rise to a complementary problem: they let in too much.

The following familiar definitions illustrate the problem: x is a circle if and only if x is a closed plane figure every point on which is equidistant from a common point; x is a circle if and only if x is a closed plane figure every arc of which has equal curvature. There seems to be nothing to recommend one over the other; if either is a correct definition, both are. In that case it seems that the following would be true by definition plus logic: x is a closed plane figure every point on which is equidistant from a common point if and only if every arc of x has equal curvature. But Kant would deem this biconditional ‘ampliative’: in any standard axiomatic formulation of geometry, the proof of it would require axioms and axioms - as opposed to definitions - are supposed to be synthetic. Evidently, this argument can be adapted to many other a priori necessities traditionally thought to be paradigmatically synthetic.

A related kind of problem arises in connection with conceptual analysis. One of the most celebrated conceptual analyses in mathematical philosophy is the classical analysis of effective calculability, or computability. On Church’s version, a function is effectively calculable if and only if it is lambda-calculable. On Turing’s version, a function is effectively calculable if and only if it is Turing computable (see [Church’s thesis](#)). Most philosophers deem each version to be a successful conceptual analysis. But when the two analyses are combined, it follows immediately that a function is recursive if and only if it is Turing computable. But this biconditional - which is an important ‘ampliative’ theorem of formal number theory - would then turn out to be analytic (in the sense of being true by conceptual analysis plus logic) even in the event that logicism is false.

These problems suggest that there is no coherent way to draw an analytic/synthetic distinction along the lines Kant thought. One response is to settle for a severely restricted use of ‘analytic’ which concerns only concepts with unique structures (that is, unique ‘decompositions’). The price of this move, however, is high: the vast majority of our concepts - including nearly all of the concepts philosophers have sought to define or analyse (good, true, valid, number, meaning, knowledge, and so on) - are in this sense unstructured and so would not give rise to new analyticities. At best, rather uninteresting concepts (such as bachelor) are of this sort. In consequence, even if knowledge of analyticities (in the narrow sense) had a transparent epistemic explanation as Kant assumed, the sort of knowledge one seeks in typical philosophical definitions or analyses would need quite another sort of explanation.

5 Epistemological problems with ‘analytic’

Kant and his successors simply assumed that knowledge of analyticities has a transparent sort of explanation (often linked to a simplistic ‘pictorial’ or ‘mereological’ view of concepts). Just what that explanation is supposed to be has never been satisfactorily stated. It cannot be that analytic propositions are those whose truth is recognized just by virtue of possessing the constituent concepts, for *no* proposition is like this. For example, it is in principle possible that someone who possesses the relevant concepts but who is in sufficiently defective cognitive conditions (deficient intelligence, attentiveness, and so on) might fail to recognize that all and only bachelors are unmarried men. It does no good to relax this account by holding that analytic propositions are those whose truth would be recognized by anyone in sufficiently good cognitive conditions just by virtue of possessing the constituent concepts, for this lets in too much: anyone in sufficiently good cognitive conditions could not fail to recognize, say, that figure A is congruent with figure B if and only if B is congruent with A . But such propositions are the very paradigms of what Kant would have deemed synthetic a priori and requiring a different sort of explanation. (For these same reasons, purely epistemic definitions of ‘analytic’ are problematic.)

Another line of explanation is to liken our knowledge, say, that bachelors are men to our knowledge that unmarried men are men, or more generally, that *ABs* are *Bs* - that is, to liken this knowledge to our knowledge of a certain kind of *logical* truth. But how do we know the logical truth that *ABs* are *Bs*? Is the explanation fundamentally different from the explanation of our knowledge of other kinds of logical truths - for example, that *Bs* are *As* or *Bs*? It is hard to see why it should be. This raises the question of how we know logical truths generally. For instance, is the explanation of our (logical) knowledge that identity is a symmetric relation ($A = B$ iff $B = A$) really different from the explanation of our (nonlogical) knowledge that congruence is a symmetric relation? From a phenomenological point of view, both instances of knowing (logical and nonlogical) arise from a priori intuitions and these intuitions, phenomenologically, are not relevantly different. On this score, therefore, there is no reason to think that our a priori knowledge divides neatly into two kinds having quite different explanations.

On the contrary, there is a promising unified explanation of a priori knowledge generally, which goes roughly as follows. In every case a priori knowledge is based evidentially on a priori intuition. The evidential force of a priori intuition is to be explained in terms of a general analysis of concept possession: it is constitutive of concept possession that in suitably good cognitive conditions intuitions regarding the behaviour of the concept need to be largely correct. If in suitably good cognitive conditions one did not have such intuitions, one would not be said to possess the concept. If something like this is right, then, although they mark cogent logical and metaphysical distinctions, all the listed uses of 'analytic' - even the narrow uses - fail to mark an epistemically significant category of knowledge.

The picture that results is complicated somewhat by Kripke and Putnam's doctrine that there are essentially a posteriori necessities - for example, $\text{water} = \text{H}_2\text{O}$. Among these are some which may plausibly be deemed *scientific definitions*. If 'analytic' is used in the sense of truth by definition plus logic, where 'definition' is understood to include scientific definitions, then there would be necessities which would be both analytic and essentially a posteriori. Evidently the Kripke-Putnam doctrine applies only to 'semantically unstable' expressions - that is, expressions ('water', 'gold', 'heat', and so on) whose meaning could be different in some population of speakers whose epistemic situation is qualitatively identical to ours. These are expressions to whose meaning the external environment makes some contribution (see **Content: wide and narrow**). The above picture, however, holds straightforwardly for semantically stable expressions ('conscious', 'know', 'good', and so on) which loom large in philosophical analysis.

6 Whither 'analytic'?

At this stage, one may reasonably ask whether continued use of 'analytic' serves any purpose in philosophy. Although the term evidently lacks the epistemological significance once attributed to it, the wider use of 'analytic' in the sense of true by definition plus logic still has utility, namely, in posing an important question: Are there necessary truths (supervenience principles, the incompatibility of colours, and so on) which are not analytic in this sense? The answer appears to be affirmative if 'definition' is understood straightforwardly as ordinary explicit definition. But if, as some have proposed, 'definition' is understood to include 'implicit definitions', the answer is controversial and depends on what information may legitimately be loaded into 'implicit definitions'. On pain of trivializing significant traditional questions, however, surely not everything is admissible. Plainly there still are unanswered questions here. But they are really about the nature of definitions; 'analytic' does no work.

See also: **Carnap, R.; Concepts; Conceptual analysis; Intensional entities; Kant, I.; Logical positivism; Necessary truth and convention**

GEORGE BEALER

References and further reading

- Ayer, A.J.** (1936) *Language, Truth and Logic*, London: Gollancz. (Uses of 'analytic' mentioned in §1.)
- Bealer, G.** (1982) *Quality and Concept*, Oxford: Clarendon Press. (Definition of 'analytic' in terms of propositions and concepts discussed in §3.)
- Bealer, G.** (1996) 'A Priori Knowledge and the Scope of Philosophy', *Philosophical Studies* 81: 121-42. (Account of a priori justification and knowledge in terms of intuition and concept possession described in §§2 and 5.)
- Boghossian, P.** (1997, forthcoming) 'Analyticity', in C. Wright and B. Hale (eds) *A Companion to the Philosophy*

- of Language*, Oxford: Blackwell.(Epistemic account of analyticity.)
- Bolzano, B.** (1837) *Wissenschaftslehre*, Sulzbach: Seidel, 4 vols; trans. and ed. R. George, in *Theory of Science*, Oxford: Blackwell, 1972; trans. B. Terrell, ed. J. Berg, in *Theory of Science*, Dordrecht: Reidel, 1973.(Use of 'analytic' mentioned in §1. The George and Terrell translations provide selections from Bolzano's work.)
- Carnap, R.** (1937) *Logical Syntax of Language*, London: Routledge.(Uses of 'analytic' mentioned in §1.)
- Carnap, R.** (1947) *Meaning and Necessity*, Chicago, IL: University of Chicago Press. (Uses of 'analytic' mentioned in §1.)
- Church, A.** (1951) 'A Formulation of the Logic of Sense and Denotation', in P. Henle, H.H. Kallen and S.K. Langer (eds) *Structure, Method and Meaning: Essays in Honor of Henry M. Scheffer*, New York: Liberal Arts Press, 3-24.(Offers non-circular definitions of modal notions.)
- Feigl, H.** (1949) 'Logical Empiricism', in H. Feigl and W. Sellars (eds) *Readings in Philosophical Analysis*, New York: Appleton-Century-Crofts, 3-26.(Use of 'analytic' mentioned in §1.)
- Fine, K.** (1994) 'Essence and Modality', *Philosophical Perspectives* 8: 1-16.(Account of analyticity and definition in terms of the notion of essence.)
- Frege, G.** (1884) *Die Grundlagen der Arithmetik: eine logisch-mathematische Untersuchung über den Begriff der Zahl*, Breslau: Koebner; trans. J.L. Austin, *The Foundations of Arithmetic: A Logico-Mathematical Enquiry into the Concept of Number*, Oxford: Blackwell, 2nd edn, 1980.(Use of 'analytic' mentioned in §1.)
- Grice, H.P.** (1989) *Studies in the Way of Words*, Cambridge, MA: Harvard University Press.(Analysis of meaning in terms of propositional attitudes discussed in §3.)
- Kant, I.** (1781/1787) *Critique of Pure Reason*, trans. N. Kemp Smith, New York: St Martin's Press, 1965. (Provides the original definition of 'analytic' and 'synthetic' and an account of the possibility of synthetic a priori knowledge.)
- Katz, J.J.** (1986) *Cogitations*, Oxford: Oxford University Press.(Account of analyticity in terms of sense structure mentioned in §§3 and 5.)
- Lewis, C.I.** (1944) 'The Modes of Meaning', *Philosophy and Phenomenological Research* 4: 236-49; repr. L. Linsky (ed.) *Semantics and the Philosophy of Language*, Urbana, IL: University of Illinois Press, 1952, 50-63. (Uses of 'analytic' mentioned in §1.)
- Lewis, D.K.** (1969) *Convention*, Cambridge, MA: Harvard University Press.(Use of 'analytic' mentioned in §1.)
- Pap, A.** (1958) *Semantics and Necessary Truth*, New Haven, CT: Yale University Press.(Use of 'analytic' mentioned in §1 and extended discussion of analyticity and related issues.)
- Parsons, C.** (1997, forthcoming) *Quine and Gödel on Analyticity*, in P. Leonardi and M. Santambrogio (eds) *On Quine*, Cambridge: Cambridge University Press.(Discussion of Gödel's uses of 'analytic' in relation to Quine.)
- Prior, A.N. and Fine, K.** (1977) *Worlds, Times and Selves*, London: Duckworth. (Defends thesis that modal notions are primitive.)
- Putnam, H.** (1962) 'The Analytic and the Synthetic', in H. Feigl and G. Maxwell (eds) *Minnesota Studies in the Philosophy of Science*, vol. 3, Minneapolis, MN: University of Minnesota Press; repr. H. Putnam, *Mind, Language and Reality: Philosophical Papers*, vol. 2, Cambridge: Cambridge University Press, 1975, 33-69. (Use of 'analytic' mentioned in §1.)
- Quine, W.V.** (1951) 'Two Dogmas of Empiricism', *Philosophical Review* 60: 20-43; repr. in *From a Logical Point of View*, Cambridge, MA: Harvard University Press, 1953, 20-46. (Classic attack on the analytic/synthetic distinction.)
- Quine, W.V.** (1960) *Word and Object*, Cambridge, MA: MIT Press.(Defence of indeterminacy of translation and attack on the ontology of propositions discussed in §3.)
- Quine, W.V.** (1970) *Philosophy of Logic*, Englewood Cliffs, NJ: Prentice Hall.(Attack on the ontology of propositions discussed in §3.)
- Schlick, M.** (1930-1) *Wissenschaftlicher Jahresbericht der Philosophischen Gesellschaft an der Universität zu Wien für das Vereinsjahr 1930/1*, repr. as 'Is There a Factual A Priori?', in H. Feigl and W. Sellars (eds) *Readings in Philosophical Analysis*, New York: Appleton-Century-Crofts, 1949, 277-94.(Use of 'analytic' mentioned in §1.)
- Strawson, P.F. and Grice, H.P.** (1956) 'In Defense of a Dogma', *Philosophical Review* 65:141-58; repr. in H.P. Grice, *Studies in the Way of Words*, Cambridge, MA: Harvard University Press, 1989, 196-212.(Early response to Quine mentioned in §1.)
- White, M.** (1950) *The Analytic and the Synthetic: An Untenable Dualism*, in S. Hook (ed.) *John Dewey: Philosopher of Science and Freedom*, New York: The Dial Press, 316-30; repr. in L. Linsky (ed.) *Semantics and the Philosophy of Language*, Urbana, IL: University of Illinois Press, 1951, 272-86.(Classic attack on the analytic/synthetic distinction.)

Antirealism in the philosophy of mathematics

Realism in the philosophy of mathematics is the position that takes mathematics at face value. According to realists, mathematics is the science of mathematical objects (numbers, sets, lines and so on); mathematicians, to use the old metaphor, are discoverers, not inventors. Moreover, just as there may be truths about physical reality which we can never know, so too, realists say, there may be truths about mathematical reality which we can never know.

It is this claim in particular which antirealists find unacceptable. Equating what can be known in mathematics with what can be proved, they insist that only what can be proved is true. (Only what can be proved: different accounts of what this 'can' means, facing different difficulties, generate different positions.) This leads antirealists to recoil not only from realism but also from the practice of mathematicians themselves. For the orthodox assumption that every mathematical statement is either true or false would be invalidated, on the antirealist view, by a statement that was neither provable nor disprovable. Not that antirealists themselves can see it in these terms. For if a statement were neither provable nor disprovable, that would itself be an unprovable truth about mathematical reality. Antirealists must learn how to be circumspect even in defence of their own circumspection.

1 Realism and its rivals

Consider the following view. Any mathematical statement such as ' $7 + 5 = 12$ ' or ' $7 + 5 = 13$ ' or 'Every even number greater than 2 is the sum of two primes' is straightforwardly true or false. If mathematical reality is as the statement says it is, then the statement is true. Otherwise, it is false. How the statement connects with mathematical reality is indicated by its grammatical structure. To each singular term there corresponds an object, to each sortal noun a kind of object, to each quantifier a range of objects, and so on. Which grammatical category an expression belongs to depends solely on how it functions. Thus '7', which functions as a singular term, is a singular term: it stands for a particular mathematical object. 'Prime', which functions as a sortal noun, is a sortal noun: it picks out a particular kind of mathematical object. 'Every even number', which functions as a quantifier, is a quantifier: it ranges over a particular domain of mathematical objects. These objects, which exist outside space and time, are mind-independent. How they stand in relation to one another is not determined by how we think they stand in relation to one another. In some cases how they stand in relation to one another is completely unknown. It may even be unknowable.

This is, in its most extreme form, 'realism' (see [Realism in the philosophy of mathematics](#)). It is also known as 'Platonism' (see [Plato §10](#)). But if this is realism, then any number of things might reasonably be meant by 'antirealism'. Each part of this view can be challenged, and has been. Thus there are those who have claimed:

- (1) That no mathematical statement is strictly true or false: mathematics consists in the manipulation of meaningless symbols in accordance with certain formal rules.
- (2) That this is true of certain auxiliary parts of mathematics: not *all* mathematical statements are strictly true or false (see [Hilbert's programme and formalism](#)).
- (3) That all mathematical statements are strictly true or false, but not straightforwardly so: ' $7 + 5 = 12$ ', for example, is like 'Hamlet was the prince of Denmark' - a fictional truth that is literally false.
- (4) That the apparent grammatical structure of mathematical statements is misleading: '7' is no more a singular term than 'John's sake' in 'She did it for John's sake'.
- (5) That the apparent grammatical structure is not misleading, but does not determine in the standard way how mathematical statements connect with mathematical reality: '7' does not stand for a particular object, but rather indicates part of a structure which any objects whatsoever could instantiate.
- (6) That there are particular mathematical objects in just the way that the grammatical structure suggests, but they are mind-dependent creations of ours: how they stand in relation to one another is not totally independent of how we think they stand in relation to one another (see [Intuitionism](#); [Constructivism in mathematics](#)). This allows for the possibility that there are certain mathematical statements, rather like 'Father Christmas is left-handed', whose truth or falsity has never been settled (see [Cantor's theorem](#); [Continuum hypothesis](#); [Existence](#)).
- (7) Finally, that how things are in mathematical reality, though it is sometimes unknown, is never unknowable. This brings us full circle, for, as we shall see, this last challenge in turn becomes a challenge to the idea that every mathematical statement is true or false.

2 The antirealist challenge

The term ‘antirealism’ is most frequently reserved for the last of the above challenges, and that is what this entry will focus on. Not that only one view about the nature of mathematical truth is in question. On the contrary, as we shall see, there are radically opposed views which issue in this particular challenge: that is, there are radically opposed reasons for thinking that mathematical truth can never outstrip knowability.

Knowability in a mathematical context can be regarded as provability. The idea of a proof must then be suitably construed. For instance, a self-evident axiom must be allowed to count as a one-line proof of itself: otherwise it would be unprovable, and hence, *ex hypothesi*, unknowable. On the other hand, clearly not every mathematical statement can count as a one-line proof of itself. Let us assume that a satisfactory account of proof can be given. The question now is: can there be unprovable mathematical truths?

One reason for denying that there can be was mentioned in the previous section. If mathematical objects are mind-dependent creations of ours, then there is nothing to them beyond what is invested in the various principles of proof that constitute the creation. The best-known exponent of such a view was L.E.J. Brouwer (1912). He held further that the raw material for the creation is provided by something essentially private and incommunicable: our experience of time. This is the position known as ‘intuitionism’.

It is a consequence of intuitionism that the meaning of a mathematical statement is itself something essentially private and incommunicable. A diametrically opposed outlook is canvassed in the writings of Michael Dummett. According to this outlook, the meaning of a mathematical statement is something essentially public and communicable. So there can be nothing more to that meaning, ultimately, than how the statement and its constituents are used in actual mathematical practice: nothing else of potential relevance can be communicated. But this too leads to a denial of the possibility of unprovable mathematical truths. For if a mathematical statement were unprovably true, then its meaning would be something that went beyond how the statement and its constituents were used, or indeed could be used.

This second brand of antirealism, which I shall refer to as ‘Dummettian’ antirealism (though Dummett has never actually endorsed it; see Dummett 1973; Dummett, M.A.E. §3), depends on broadly Wittgensteinian considerations about the nature of meaning (see Wittgenstein, L. §12). Such considerations could just as well be applied in a non-mathematical context. Nevertheless, there are reasons for thinking that they are peculiarly suited to mathematics. For whether or not mathematical objects are mind-dependent creations of ours, there is surely something in the idea that they are nothing apart from the techniques and proof-procedures of mathematical practice. They lack the ‘solidity’ of physical objects. So even if there is some prospect that we can capture such ‘solidity’ in our use of language, and thereby express inaccessible truths about physical objects, there is no prospect that we can do anything similar in the case of mathematical objects.

A third brand of antirealism takes off from the observation that mathematics makes essential use of infinite domains. On one very attractive view, an infinite domain is unlike a finite domain in that its members cannot be given ‘all at once’. Somewhat less metaphorically, they cannot all have a separate involvement in any one state of affairs, or make a separate contribution to any one truth. For a statement about them all to be true, there has to be a single principle which determines its truth. That is, there has to be a proof.

All these brands of antirealism have a revisionary impact on standard mathematical practice. Consider a statement that has never been proved or disproved. At the time of writing, Goldbach’s conjecture - that every even number greater than 2 is the sum of two primes - is an example. It is part of standard mathematical practice to assume that the statement is either true or false. (This is an application of the law of the excluded middle.) Antirealists cannot share this assumption. To make the assumption is to register indifference about whether the statement can be proved or disproved. It is to allow that the statement may be true without proof, a kind of ‘infinite coincidence’. This is precisely what antirealists cannot allow.

3 What kind of possibility does antirealism involve?

The antirealist view is that only what can be proved is true. A much more radical view (not without adherents) is that only what *has been* proved is true. By holding back from this more radical view, antirealists inevitably draw

attention to the ‘can’. But what sense of ‘can’ is intended? Certainly, it is what can be proved *by us* that counts, and not, for example, by an infinite being existing outside space and time. But that does not answer the question.

Basically, antirealists need to decide whether they mean ‘can in principle’ or ‘can in practice’. Not that a simple one-or-other verdict will do. Both ‘can in principle’ and ‘can in practice’ require further elucidation. (How and where do scientific limitations figure? Or technological limitations? Or medical limitations?) But the broad distinction remains. And it is clear that if antirealists mean ‘can in practice’, then the revisionary impact on standard mathematics is going to be enormous. There are some numbers which are so big that we cannot, in practice, tell whether or not they are prime. So given such a number, we shall not be entitled to assume that the number is either prime or not. More generally, since mathematics cannot be extended to the unmanageably big, we shall not be entitled to assume that mathematical reality is infinite.

The resultant version of antirealism is known as ‘strict finitism’. Many antirealists think that strict finitism is incoherent. Call such antirealists ‘moderates’. For them, the ‘can’ must mean ‘can in principle’. But they need to explain why, given that their arguments would break down if the ‘can’ did not mean that, the arguments are not to be distrusted anyway. Dummettian moderates in particular are likely to experience some difficulty here. Others, whose very starting point is that mathematical reality is infinite, experience no difficulty at all.

A quite different problem for moderate antirealists is to explain why their own version of antirealism is not just a terminological variant of realism. For cannot any statement be proved or disproved *in principle*? Reconsider Goldbach’s conjecture. What is wrong with the following recipe for either proving it or disproving it? ‘Check successive even numbers greater than 2. In each case, ascertain whether or not it is the sum of two primes. Keep going until either every such number has been checked or a counterexample to the conjecture has been found.’ It is of no avail to reply simply that this process might never end. Something needs to be said to forestall the objection that if we spend half a minute checking 4, a quarter of a minute checking 6, an eighth of a minute checking 8, and so on, then the process will end in a minute – at most. Of course, we cannot in practice do this. But that is beside the point.

It is by no means clear that moderate antirealists have nothing to say here. But they may have nothing to say which is not question-begging. For instance, they may be forced to deploy a conception of what is possible in principle which itself incorporates an antirealist understanding of what it is to do infinitely many different things. If so, that is not an objection to moderate antirealism. But it does raise the spectre of realists and moderate antirealists operating with two quite different, incommensurable, individually coherent conceptions of what is possible in principle, together with two correspondingly different conceptions of how to do mathematics. Ultimately, perhaps, the realist is prepared to make certain assumptions in mathematics which the moderate antirealist is not prepared to make, and, to the extent that they are incapable of finding any independent leverage to settle their differences, they are talking past each other.

4 Implications for mathematical practice

The sheer fact that realists are willing to endorse standard mathematical practice might be thought to count in their favour. For, arguably, reversing Marx’s dictum, it is not the business of philosophers of mathematics to *change* mathematics, only to interpret it. (Wittgenstein famously wrote that ‘philosophy leaves mathematics as it is’.) Here, as before, there is a particular difficulty for Dummettian antirealists. For they are only too keen to cast actual mathematical practice in the role of a datum. Dummett’s own reaction to this difficulty is to argue that, even if actual mathematical practice is a datum, it is not an indissoluble datum: nothing in the view he canvasses precludes criticizing some parts of that practice for failing to harmonize with other parts.

Suppose there *is* room for a non-conservative philosophy of mathematics. Antirealists face a further difficulty. They have no satisfactory answer to the question ‘When exactly does standard mathematical practice lead us astray?’. They cannot answer, ‘Whenever a statement is neither provable nor disprovable’. For if a statement were neither provable nor disprovable, that would itself be an unprovable truth about mathematical reality. (To prove that a statement was not provable would already be to disprove it.) Likewise, if, say, there were no proof of Goldbach’s conjecture, nor, as it happened, a counterexample, that would itself be an ‘infinite coincidence’ of the very kind that antirealists disavow. Antirealists have no option, then, but to adopt a kind of stoic silence. Whenever others make assumptions in mathematics which they are not themselves prepared to make, they must

withhold assent, but they have no satisfactory way of saying what is holding them back. As far as anything they can say is concerned, their restraint may be a result, not of nonconformity, but of sheer reticence.

Antirealism remains of the utmost importance, despite this very real difficulty. Indeed, part of its importance may be in drawing the difficulty to our attention. For if antirealists are right, then it may be that their knowledge of correct mathematical practice is a prime example of a fundamental and neglected philosophical category: knowledge that is essentially ineffable.

See also: [Logical and mathematical terms, glossary of](#)

A.W. MOORE

References and further reading

- Benacerraf, P. and Putnam, H.** (eds) (1964) Introduction to *Philosophy of Mathematics: Selected Readings*, Cambridge: Cambridge University Press, 2nd edn, 1983, 1-37.(Useful survey of the basic issues in the philosophy of mathematics, touching helpfully on both realism and antirealism.)
- Bernays, P.** (1935) 'Sur le Platonisme dans les mathématiques', *L'enseignement mathématique*, 1st series, 34: 52-69; trans. C.D. Parsons, 'On Platonism in Mathematics', in P. Benacerraf and H. Putnam (eds) *Philosophy of Mathematics: Selected Readings*, Cambridge: Cambridge University Press, 2nd edn, 1983, 258-71.(Classic survey of Platonism/realism and standard antirealist reactions.)
- Brouwer, L.E.J.** (1912) *Intuitionisme et Formalisme*, Groningen: Noordhoff; trans. A. Dresden, 'Intuitionism and Formalism', *Bulletin of the American Mathematical Society* 20: 81-96, 1913; repr. in P. Benacerraf and H. Putnam (eds) *Philosophy of Mathematics: Selected Readings*, Cambridge: Cambridge University Press, 2nd edn, 1983, 77-89. (Classic statement of some of the fundamental tenets of intuitionism.)
- Dummett, M.A.E.** (1973) 'The Philosophical Basis of Intuitionist Logic', in H.E. Rose and J.C. Shepherdson (eds) *Proceedings of the Logic Colloquium, Bristol, July 1973*, Amsterdam: North Holland, 1975, 5- 40; repr. in *Truth and Other Enigmas*, London: Duckworth, 1978, 215-47; and in P. Benacerraf and H. Putnam (eds) *Philosophy of Mathematics: Selected Readings*, Cambridge: Cambridge University Press, 2nd edn, 1983, 97-129.(Expounds the argument sketched in §2 for 'Dummettian' antirealism.)
- Dummett, M.A.E.** (1977) *Elements of Intuitionism*, Oxford: Clarendon Press, 1990.(A very thorough account of intuitionistic revisions to mathematics, together with a full discussion of antirealist reasons for adopting them.)
- George, A.** (1993) 'How Not to Refute Realism', *Journal of Philosophy* 90 (2): 53-72.(Identifies question begging in certain antirealist arguments, and lends support to the suggestion of §3 that realism and antirealism may be two equally acceptable but incommensurable alternatives.)
- Lear, J.** (1977) 'Sets and Semantics', *Journal of Philosophy* 74 (2): 86-102.(Argues that even realists have reasons for adopting intuitionistic revisions to mathematics.)
- McDowell, J.** (1989) 'Mathematical Platonism and Dummettian Anti-Realism', *Dialectica* 43 (1/2): 173-92. (Endorses the suggestion of §2 that the basic considerations in favour of 'Dummettian' antirealism apply only in a mathematical context, where the objects of investigation lack 'solidity'.)
- Prawitz, D.** (1977) 'Meaning and Proofs: On the Conflict Between Classical and Intuitionistic Logic', *Theoria* 43 (1): 2-40.(Good general account of the argument expounded in Dummett (1973) and sketched in §2).
- Putnam, H.** (1980) 'Models and Reality', *Journal of Symbolic Logic* 45 (3): 464-82; repr. in P. Benacerraf and H. Putnam (eds) *Philosophy of Mathematics: Selected Readings*, Cambridge: Cambridge University Press, 2nd edn, 1983, 421-44.(Uses the Löwenheim-Skolem theorem to attack extreme realism.)
- Rotman, B.** (1993) *Ad Infinitum... The Ghost in Turing's Machine*, Stanford, CA: Stanford University Press. (Defends a version of strict finitism.)
- Wittgenstein, L.J.J.** (1956) *Remarks on the Foundations of Mathematics*, ed. G.H. von Wright, R. Rhees and G.E.M. Anscombe, trans. G.E.M. Anscombe, Oxford: Blackwell, 2nd edn, 1967; 3rd edn, 1978.(*Locus classicus* of Wittgenstein's later philosophy of mathematics, with its own distinctive brand of antirealism.)
- Wright, C.** (1980) *Wittgenstein on the Foundations of Mathematics*, London: Duckworth.(Very thorough discussion of antirealism in Wittgenstein's later philosophy of mathematics.)
- Wright, C.** (1982) 'Strict Finitism', *Synthese* 51: 203-82; repr. in *Realism, Meaning and Truth*, Oxford: Blackwell, 1986; and in A.W. Moore (ed.) *Infinity*, Aldershot: Dartmouth, 1993, 251-330.(Sympathetic discussion of strict finitism and its revisionary impact on mathematics.)

Arithmetic, philosophical issues in

The philosophy of arithmetic gains its special character from issues arising out of the status of the principle of mathematical induction. Indeed, it is just at the point where proof by induction enters that arithmetic stops being trivial. The propositions of elementary arithmetic - quantifier-free sentences such as ' $7 + 5 = 12$ ' - can be decided mechanically: once we know the rules for calculating, it is hard to see what mathematical interest can remain. As soon as we allow sentences with one universal quantifier, however - sentences of the form ' $(\forall x)f(x) = 0$ ' - we have no decision procedure either in principle or in practice, and can state some of the most profound and difficult problems in mathematics. (Goldbach's conjecture that every even number greater than 2 is the sum of two primes, formulated in 1742 and still unsolved, is of this type.)

It seems natural to regard as part of what we mean by natural numbers that they should obey the principle of induction. But this exhibits a form of circularity known as 'impredicativity': the statement of the principle involves quantification over properties of numbers, but to understand this quantification we must assume a prior grasp of the number concept, which it was our intention to define. It is nowadays a commonplace to draw a distinction between impredicative definitions, which are illegitimate, and impredicative specifications, which are not. The conclusion we should draw in this case is that the principle of induction on its own does not provide a non-circular route to an understanding of the natural number concept. We therefore need an independent argument. Four broad strategies have been attempted, which we shall consider in turn.

1 Formalism

Although the problem of accounting for both the objects apparently spoken of and the knowledge we appear to have about them in a way that explains both the apparent necessity and the apparent applicability of this knowledge is one which arises for all areas of mathematics, arithmetic has features which make some answers to the question seem especially plausible. In the language we use to express arithmetic every number has a canonical name (called a 'numeral'). This fact has made formalism a persistently attractive option in the case of arithmetic: since every number has a numeral to represent it, we can restrict our attention to the symbols and let numbers themselves drop out of the account entirely. It is not clear that this move helps very much with the epistemological part of the problem, since if the account is to license the whole of arithmetic, numerals will have to be abstract entities our knowledge about which demands explanation. But in any case formalism as just described is untenable, since it leaves the applicability of arithmetic wholly unexplained: we cannot rest content with the claim that arithmetic is just a game with symbols and not demand an account of why it should be *this* game that we play rather than any other.

If the fault in formalism is that it does not allow arithmetical statements to be meaningful, it might be thought an improvement to regard the terms occurring in our formal theory as being given their meaning by the role they play in the axioms of the theory. This variant of formalism has been advanced as an account not just of Peano's axiomatization of arithmetic but of the axiomatic method generally (most famously by Hilbert in his correspondence with Frege; see Frege 1980). It is far from clear, however, that the terms of a formal language can be invested with meaning this easily. In the 1920s and 1930s the members of the Vienna Circle (most notably their 'logician-in-chief', Rudolf Carnap) regarded arithmetic as true only 'by convention'. But if we were to adopt an inconsistent convention our language would be incapable of saying anything about the world. The general point, then, is that formalism cannot explain applicability without assuming consistency, and the most obvious sort of proof of the consistency of arithmetic appeals to our prior grasp of an infinite model, thus facing again the very difficulty formalism was intended to solve.

In his later work Hilbert had the idea of partitioning arithmetic into a real part, for which he gave a finitistic justification (see §3 below), and an ideal part, to be treated formally and justified instrumentally as a method of providing shorter or more comprehensible proofs of results in the real part. The consistency problem arises once more, of course, but Hilbert hoped to be able to solve it without supplying a model, by a combinatorial analysis of the syntax of the formal language. Such an analysis could, he thought, be carried out using the conceptual resources of the finitistically justified real part. Gödel's incompleteness theorems coupled with Turing's analysis of the notion of computability killed Hilbert's programme in the grand form originally envisaged. However,

mathematicians hardly ever exploit the full strength of the classical system and not everyone has given up hope of an instrumental justification on Hilbertian lines of the parts they actually use (see [Gödel's theorems §6](#); [Hilbert's programme and formalism §4](#); [Turing, A.M.](#)).

2 Empiricism

If objections to formalism lead us to believe that arithmetical statements really are *true*, we might try Mill's strategy, which treats numbers as empirical and arithmetical statements as empirical generalizations (see [Mill, J.S. §3](#)). This strategy has not been popular - our conviction that the hardness of the logical 'must' applies in arithmetic too seems difficult to dispel - but it is quite difficult to refute decisively. Frege objected to it on the grounds that if numbers were empirical, only what is empirical could be counted, whereas number, in Locke's phrase, 'applies itself to men, angels, actions, thoughts - everything that either doth exist or can be imagined' (1689: II.xvi.1). This is not, however, a decisive objection, since the empiricist may be prepared to accept that 'everything that doth exist or can be imagined' is ultimately explicable in empirical terms. A better objection is that empiricism seems ill-equipped to ground our knowledge about large numbers (for example, those larger than the number of atoms in the observable universe). Even though such numbers are practically useless (which is precisely why empirical evidence provides no confirmation for assertions about them), the revision to arithmetic which would be needed if we did not admit them is radical and the resulting system would be awkward to use. The most persuasive objection of all, however, is the difficulty the arithmetical empiricist has in justifying the principle of mathematical induction. It is very hard to see what would count as empirical evidence either for or against this principle.

3 Intuitionism

A third broad strategy is to appeal to intuition to explain how arithmetical knowledge is possible. This approach is due to Kant, whose account of arithmetic parallels quite closely his account of geometry. A geometric proof typically begins with a construction (of a triangle, for instance), on which various operations are then performed. Since I construct the triangle in intuition, Kant held that it will be subject to just the spatial structure which sensibility imposes on all my experience. As a consequence, the geometric theorem I prove by this means is applicable a priori to experience. Similarly, in order to prove that $7 + 5 = 12$ I must first construct the concept 'seven' in intuition and then perform on my construction various operations. What is less clear in the arithmetical case than in the geometric one is how the structure imposed by sensibility - what Kant calls the 'form' of intuition - constrains the construction. Kant admits that we do not depend on sensibility for the *concepts* of arithmetic, which are therefore purely intellectual in contrast to those of geometry, but he holds that constructing the concepts depends on counting, which is a process apparently subject to the forms of space and time. It was common in the nineteenth century to read Kant as advocating a neat parallelism - that arithmetic is the science of time just as geometry is that of space - but this is at best a simplification. It is in any case far from clear that counting depends on time. Notice, though, that if we abandon this dependence we thereby give up just what Kant saw as the principal advantage of his account, namely its explanation of the applicability of arithmetic. The moral is quite general: the appeal to intuition beyond the limits of logic, which is characteristic of Platonism, does nothing to explain the applicability of mathematics when it is removed from the Kantian framework (see [Kant, I. §5](#)).

If I assert an arithmetical proposition, I am, according to Kant, reporting the result of a mental construction which I have performed or could in principle perform. But as Frege said, 'another man's idea is another idea'. Kant's account makes ' $7 + 5 = 12$ ' express a different statement on each person's lips. It is therefore unavoidably solipsistic.

A further difficulty is caused by Kant's insistence that sensibility is passive and that the intuitions it supplies are immediate representations: it is hard to believe that I can have in my mind an *immediate* representation of a very large number, such as 10^{100} . The notion (denied by Kant) that humans are capable not only of passive intuition through the faculty of sensibility but also of creative - Kant calls them 'intellectual' - intuitions is a theme of German romanticism in the nineteenth century. In application to arithmetic it was used by Brouwer to provide an account on which our experience of time unfolding gives us an intuition of succession - what Brouwer calls the bare two-oneness - which we can repeat for ourselves at will as creative subjects. On this view the law of the excluded middle is not justified: if numbers are mind-dependent and have no existence outside of our constructions, then, for example, there is no reason to believe that Goldbach's conjecture (that every even number greater than 2 is the sum of two primes) must be either true or false in advance of constructing either a proof or a

counterexample.

Hilbert's justification for the real part of his real/ideal partition mentioned in §1 above proceeds on similar lines to Brouwer's, although the intuitions on which numbers are based are for Hilbert those of finite arrangements of concrete objects. For this reason Hilbert's account seems less likely than Brouwer's to descend into solipsism. But the principal difference between their conceptions is that Brouwer regarded arithmetical statements and proofs as mental constructions available within arithmetic in the same way as numbers themselves. For Hilbert, on the other hand, any application of arithmetic to proofs is a matter for the metatheory. The result is that Hilbert's finitism can justify directly only a fragment of arithmetic with a limited induction principle (so-called primitive recursive arithmetic), whereas Brouwer's intuitionism has a much stronger induction principle available to it.

There is a persistent worry, however, as to whether either Hilbert or Brouwer overcame the problem Kant's account faced with large numbers. Both were ready to accept arguments on the basis of what humans can do in principle: these arguments seem to turn on a distinction between finite and infinite which is not available from their perspective without independent justification. If so, their position collapses into 'strict finitism', the view that we cannot have a general grasp of the notion of arbitrarily large finite numbers.

4 Logicism

The last strategy, the formulation of which is due to Frege, is to show that arithmetical truths are logical. One way of doing this is to postpone the attempt to explain the substantival use of number-words and to explain instead their adjectival use. This can be done by defining not the numbers themselves but the numerically definite quantifiers:

$$\exists^0 x Fx \equiv_{\text{Df}} \sim \exists x Fx$$

$$\exists^{n+1} x Fx \equiv_{\text{Df}} \exists x (Fx \wedge \exists^n y (Fy \wedge y \neq x)).$$

From these definitions we can prove logically that, for example,

$$\exists^7 x Fx \wedge \exists^5 x Gx \wedge \exists^0 x (Fx \wedge Gx) \supset \exists^{12} x (Fx \vee Gx),$$

which we can interpret as being what is meant by '7 + 5 = 12'. By the same means we can generate the whole of the positive part of elementary arithmetic, that is, the true quantifier-free equalities. What we cannot generate is the negative part, that is, the true inequalities. For this we would need a guarantee, which logic alone cannot supply, that there is no finite bound to the number of objects that there are. (Wittgenstein in the *Tractatus* (1922) proposed a rather more general variant of the same strategy, but it seems to be susceptible, in the context of Wittgenstein's system, to just the same difficulty over proving inequalities.)

Frege considered the introduction of numbers via numerically definite quantifiers in his *Grundlagen* (1884). What led him to reject it was not the difficulty of deriving inequalities but a problem facing not just this but any strategy involving implicit definitions, namely that such definitions do not fully determine the identity conditions of the objects they attempt to introduce. This is nowadays known as the Julius Caesar problem because Frege posed it by asking how one could tell from the implicit definitions whether or not Julius Caesar is a number. In the present context this question creates a problem not for the project of explaining the adjectival use of number-words but for the strategy of explaining their substantival use in terms of their adjectival use. The problem bites, therefore, only when we attempt to move beyond elementary arithmetic and quantify over numbers.

Another logicist strategy considered and rejected by Frege, but the subject of renewed attention at the end of the twentieth century, is to derive arithmetic from the 'numerical equivalence', that is, the contextual principle (sometimes called 'Hume's principle' or 'N=') that the number of *F*s is the same as the number of *G*s if and only if the *F*s and the *G*s can be correlated one-to-one. As a technical programme this is perfectly feasible, as Frege was the first to realize. He sketched in the *Grundlagen* a construction of the natural numbers and a proof that they satisfy Peano's axioms, assuming only second-order logic and the numerical equivalence. He nevertheless rejected the strategy of basing arithmetic on this equivalence because the Julius Caesar problem applies to it as much as to the previous contextual strategy. Interest in the strategy has been revived by Crispin Wright, who argued (1983) that the Julius Caesar problem can be solved by appeal to an independently plausible 'sortal inclusion principle' to the effect that objects must be of different sorts if the content of their identity conditions is sufficiently different.

The proposal Frege settled on instead was to define the number of *F*s explicitly as the class of all concepts

equinumerous with *F*. He showed in *Grundgesetze der Arithmetik* (1893-1903) that on this definition of number arithmetic can be deduced from what he took to be logical principles. (One of his allegedly logical principles was inconsistent, however, and subsequent attempts - most notably by Whitehead and Russell - to repair Frege's system have had to appeal to principles even their advocates have balked at calling logical.)

Frege's solution to the Julius Caesar problem (the underdetermination of the objects of arithmetic by the principles about them to which we are committed) is in any case under threat from the opposite problem: any non-arithmetical determination of the objects which solves the Julius Caesar problem will give numbers extra properties which, since they do not flow from the principles governing numbers, must be arbitrary and hence spurious. This problem was mentioned by Dedekind (in a letter to Weber dated 24 January 1888; see Ewald 1996), but it has come to prominence more recently through a much-cited paper by Benacerraf (1965). Two ways of dealing with it have been advanced under the generic label of 'structuralism'. Dedekind's way - that we are capable, once we have given a logical construction of one model of Peano's axioms, of abstracting away from its particular features to gain a conception of a model without those features - appeals to a mental process (abstraction) that many have found mysterious. Benacerraf's way - that arithmetic should be seen as the study not of one particular model of Peano's axioms but of the structure which all such models have in common - is in danger of relapsing into the axiomatic formalism we considered earlier.

What Russell's paradox has suggested to Michael Dummett is that the concept 'set' is what he calls 'indefinitely extensible' (see Dummett 1991; Dummett, M.A.E. §3). This means that any attempt to regard the objects falling under the concept as forming a definite totality leads inevitably to the realization that there are other objects not in the totality which we are nevertheless forced to admit as falling under the concept. Dummett holds that the presence of 'indefinitely extensible' concepts is a characteristic feature of mathematics which should lead us to espouse for it the anti-realism which his more general meaning-theoretic arguments make room for. He recommends that we abandon the law of the excluded middle and espouse the mathematics of intuitionism but not Brouwer's solipsistic conception of its objects. Dummett thinks that in this way we can retain Frege's logicist insight that numbers are abstract objects truths about which embody deductive subroutines whose application to the world is validated by logic alone.

Just as Brouwerian intuitionism has been accused of an instability which reduces it to strict finitism, not everyone is persuaded that Dummett's position does not collapse into the ultra-intuitionism of middle-period Wittgenstein, according to which the meaning of an arithmetical generalization is identical with its proof. On this view Goldbach's conjecture does not in the present state of knowledge have any meaning at all.

See also: Frege, G.; Husserl, E.; Logical and mathematical terms, glossary of

MICHAEL POTTER

References and further reading

- Benacerraf, P.** (1965) 'What Numbers Could Not Be', *Philosophical Review* 74: 47-73; repr. in P. Benacerraf and H. Putnam (eds) *Philosophy of Mathematics: Selected Readings*, Cambridge: Cambridge University Press, 2nd edn, 1983. (Referred to in §4. The impetus for much recent work on the philosophy of arithmetic.)
- Benacerraf, P. and Putnam, H.** (eds) (1964) *Philosophy of Mathematics: Selected Readings*, Cambridge: Cambridge University Press, 2nd edn, 1983. (Includes work by Hilbert and Benacerraf referred to in this entry, as well as several other relevant articles.)
- Carnap, R.** (1934) *Logische Syntax der Sprache*, Vienna: Springer; trans. A. Smeaton, *The Logical Syntax of Language*, London: Kegan Paul, 1937. (The most developed expression of the Vienna Circle's view of mathematics as being true by convention.)
- Dedekind, R.** (1888) *Was sind und was sollen die Zahlen?*, Braunschweig: Vieweg; trans. 'What are Numbers and What Should They Be?', in W.B. Ewald (ed.) *From Kant to Hilbert: A Source Book in the Foundations of Mathematics*, Oxford: Clarendon Press, 1996. (A beautiful presentation of arithmetic as the study of 'simply infinite' systems, that is, models of what are now known as Peano's axioms.)
- Dummett, M.A.E.** (1991) *Frege: Philosophy of Mathematics*, London: Duckworth. (A lucid critique of Frege's logicism and a brief account of Dummett's own views on indefinitely extensible concepts.)
- Ewald, W.B.** (ed.) (1996) *From Kant to Hilbert: A Source Book in the Foundations of Mathematics*, Oxford: Clarendon Press. (Includes the letter from Dedekind to Weber referred to in §4, as well as much relevant material by Kant, Hilbert, Brouwer and others.)

- Frege, G.** (1884) *Die Grundlagen der Arithmetik: eine logisch-mathematische Untersuchung über den Begriff der Zahl*, Jena: Pohle; trans. J.L. Austin, *The Foundations of Arithmetic: A Logico-Mathematical Enquiry into the Concept of Number*, Oxford: Blackwell, 2nd edn, 1980. (Referred to in §4. An accessible, entertaining classic.)
- Frege, G.** (1893-1903) *Grundgesetze der Arithmetik*, Jena: Pohle, 2 vols; Part 1 of vol. 1 trans. M. Furth, *Basic Laws of Arithmetic*, Berkeley, CA: University of California Press, 1964. (Referred to in §4. Despite the contradiction at its core, includes much that is still of interest.)
- Frege, G.** (1980) *Philosophical and Mathematical Correspondence*, trans. H. Kaal, ed. G. Gabriel *et al*, Oxford: Blackwell. (Includes the correspondence between Frege and Hilbert referred to in §1.)
- Hilbert, D.** (1926) 'Über das Unendliche', *Mathematische Annalen* 95: 161-90; trans. 'On the Infinite', in P. Benacerraf and H. Putnam (eds) *Philosophy of Mathematics: Selected Readings*, Cambridge: Cambridge University Press, 2nd edn, 1983. (A statement of Hilbert's programme.)
- Locke, J.** (1689) *An Essay concerning Human Understanding*, ed. P.H. Nidditch, Oxford: Clarendon Press, 1979. (Referred to in §2. Includes an account of arithmetic that presages some aspects of Kant's.)
- Wittgenstein, L.** (1922) *Tractatus Logico-Philosophicus*, trans. D.F. Pears and B.F. McGuinness, London: Routledge, 1961. (Referred to in §4. The treatment of arithmetic is to be found in the 6.1s and 6.3s.)
- Wittgenstein, L.** (1929-30) *Philosophical Remarks*, ed. R. Rhees, trans. R. Hargreaves and R. White, Oxford: Blackwell, 1978; repr. Chicago, IL: University of Chicago Press, 1980. (The source for the ultra-intuitionistic conception of Wittgenstein's middle period.)
- Wright, C.** (1983) *Frege's Conception of Numbers as Objects*, Aberdeen: Aberdeen University Press. (Referred to in §4. A fascinating defence of the contextual strategy Frege rejected.)
- Wright, C.** (1986) 'Strict Finitism', in *Realism, Meaning and Truth*, Oxford: Blackwell. (A rare recent example of a philosopher taking strict finitism seriously.)

Axiom of choice

The axiom of choice is a mathematical postulate about sets: for each family of non-empty sets, there exists a function selecting one member from each set in the family. If those sets have no member in common, it postulates that there is a set having exactly one element in common with each set in the family. First formulated in 1904, the axiom of choice was highly controversial among mathematicians and philosophers, because it epitomized 'non-constructive' mathematics. Nevertheless, as time passed, it had an increasingly broad range of consequences in many branches of mathematics.

1 History of the axiom

Consider a whimsical example, due to Bertrand Russell, illustrating the need for an axiom of choice (stated above). Suppose a millionaire buys as many pairs of socks as pairs of boots and suppose that at last he has infinitely many pairs of each. Then (1) he has exactly as many boots as he does pairs of boots, and (2) he has exactly as many socks as he does pairs of socks. To show (1), we make use of the fact that there is a rule to pick a boot from each pair (for example, the left one). To establish (2) we must be able to pick a sock from each pair. Leaving aside artificial devices, such as the position of socks in space, there is no way to distinguish a sock in each pair and hence no way to prove (2). The axiom of choice implies the existence of such a choice from each pair, even when no rule is available to make the choice.

Thus the axiom involves infinitely many choices from sets. Despite misunderstandings by some philosophers, it has nothing to do with a single choice from a set or even with a finite number of choices, since these are already validated by first-order logic. Nor does the axiom concern the empirical possibility of actually doing the choosing, although this red herring was once taken seriously.

When first formulated by Ernst Zermelo in 1904, the axiom of choice was highly controversial. It led a number of British, French and German mathematicians to consider the philosophical side of mathematics, and induced several of them, especially in France, to adopt various constructivist philosophies of mathematics (see [Constructivism in mathematics](#)). At issue was the question of what is meant by 'existence' in mathematics.

The axiom of choice also motivated the axiomatization of set theory. Concerned primarily to defend this axiom against its numerous critics and secondarily to avoid the contradictions recently discovered (see [Paradoxes of set and property](#)), Zermelo axiomatized set theory in 1908. The axiom of choice figured prominently in his axiomatization. It turned out that many mathematical theorems required the axiom, and quite a few were actually equivalent to it (see §3 below).

The axiom is intimately involved with the emergence of important techniques for constructing models of set theory in first-order logic. The first such technique (1938) was Kurt Gödel's constructible sets (see [Constructible universe](#)). The second (1963) was Paul Cohen's method of forcing. These techniques established that the axiom of choice is consistent with the usual axioms of set theory, but also independent from them.

2 Philosophical questions surrounding the axiom

Today almost all mathematicians accept the axiom of choice, although many still remark when they use it. Earlier, Zermelo defended the axiom against its critics on two grounds: its usefulness and its self-evidence. He remarked that it had already been widely used, without being formulated explicitly. Such widespread use, he argued, could only be explained by the axiom's self-evidence.

Opponents of the axiom had various views. Some, such as Emile Borel, accepted the axiom when applied to a denumerable family of sets, but rejected it when applied to a family of larger cardinality, such as the set of all real numbers. Others, including René Baire, tentatively accepted the real numbers but argued against the existence of the set of all subsets of the real numbers, from which subsets Zermelo's choices were made. Many opponents rejected the axiom because it provided no rule for making the choices. Apparently thinking of axioms in the traditional way as self-evident truths, they often failed to understand that the axiom is needed precisely when no rule is available.

Opponents were encouraged by the axiom's counter-intuitive consequences. The most vivid of these was the

Banach-Tarski paradox, whereby a sphere S can be decomposed into a finite number of pieces and reassembled into two spheres each of the same size as S . The three-dimensionality is important here, since such a decomposition cannot be done with a two-dimensional figure such as a circle. Euclidean geometry also plays a role, since hyperbolic geometry (in which the parallel postulate is false) allows the various counter-intuitive decompositions without using the axiom.

The irony for the axiom's opponents was that if set theory is understood in a sufficiently constructivist way, then the axiom is true. In the extreme case, when all sets are finite, the axiom is true but trivial. If set theory is restricted to Gödel's constructible sets, then again the axiom is true but now distinctly non-trivial. It is only when set theory is taken in a non-constructive way that the axiom might be false. Even in intuitionism, various restricted forms of the axiom are true, but their meaning is changed since the underlying logic is not first-order logic (see [Intuitionism](#)).

3 Mathematical theorems that need the axiom

Zermelo introduced the axiom of choice to prove the well-ordering theorem: every set can be well-ordered (see [Set theory](#)). Previously Georg Cantor had assumed this well-ordering theorem as a 'law of thought', but mathematicians had been sceptical. Zermelo showed that the well-ordering theorem follows from an axiom that is conceptually much simpler: the axiom of choice.

The axiom of choice is vital to the arithmetic of infinite cardinal numbers. In particular, it is equivalent to the proposition that if two cardinals are unequal, then one is greater than the other. It is needed in order to define the addition or multiplication of infinitely many cardinals.

Various definitions of the finiteness of a set involve the axiom. The usual definition says that a set M is finite if it is empty or if there is some counting number n such that M can be put into one-one correspondence with $\{1, 2, \dots, n\}$; otherwise M is infinite. Richard Dedekind formulated the first definition that made no use of numbers. A set M is said to be 'Dedekind-infinite' if there is a one-one function from M to a proper subset of itself; otherwise, M is 'Dedekind-finite'. The philosophical significance of this definition is not only that it dispensed with numbers but also that the infinite is defined positively, rather than as the absence of some property. Intuitively, a set should be finite if and only if it is Dedekind-finite, but this equivalence requires a weak form of the axiom of choice. Moreover, there are definitions of finiteness whose equivalence to the usual definition is itself equivalent to the axiom.

The axiom permeates most branches of mathematics, except geometry. Without the axiom, much of algebra, analysis and topology would not exist. A typical algebraic theorem equivalent to the axiom is that every vector space has a basis. In topology, an equivalent theorem is the proposition that the product of compact topological spaces is compact. In analysis, we rarely need the full strength of the axiom but can get along with a weaker assumption: the principle of dependent choices. This principle asserts the existence of a sequence of choices, each depending on the finitely many choices previously made.

In logic, the axiom of choice also plays an essential role in many important results. It is needed to prove the completeness theorem for first-order logic, as well as the compactness theorem and the completeness theorem for propositional logic, provided that the set of primitive symbols of the logic may have any infinite cardinality. With such a language, the downward Löwenheim-Skolem theorem is also equivalent to the axiom: every model of an infinite set M of sentences has a model whose cardinal is at most that of M (see [Löwenheim-Skolem theorems and nonstandard models](#); [Model theory](#)). It is ironic, and of philosophical significance, that the axiom must be used to prove, via forcing, that there exists a model of set theory in which the axiom is false.

The axiom of choice is implied by various stronger propositions in set theory, such as the axiom of constructibility and the generalized continuum hypothesis (see [Constructible universe](#); [Continuum hypothesis](#)). Another such proposition, which, unlike those two, is very intuitive, was discovered by John von Neumann: a class M is a proper class (that is, not a set) if and only if M can be put into one-one correspondence with the class of all sets. This beautiful proposition turns Burali-Forti's paradox (see [Paradoxes of set and property](#)) into an argument that the universe of all sets is well-ordered, and illustrates how paradoxical results, if viewed correctly, can often be transformed into useful theorems.

See also: [Logical and mathematical terms, glossary of](#)

GREGORY H. MOORE

References and further reading

- Jech, T.J.** (1973) *The Axiom of Choice*, Amsterdam: North Holland.(A careful study of the independence results surrounding the axiom.)
- Moore, G.H.** (1982) *Zermelo's Axiom of Choice: Its Origins, Development, and Influence*, New York: Springer.(A detailed historical treatment of the axiom from its prehistory to the independence proofs. An appendix gives a translation of the most important document by the French opponents of the axiom.)
- Rubin, H. and Rubin, J.E.** (1985) *Equivalents to the Axiom of Choice II*, Amsterdam: North Holland. (Statements of many propositions equivalent to the axiom, with proofs.)
- Russell, B.A.W.** (1919) 'Selections and the Multiplicative Axiom', in *Introduction to Mathematical Philosophy*, London: Allen & Unwin; repr. London: Routledge, 1993.(An introductory discussion of the axiom of choice, which Russell calls the multiplicative axiom, by one of its most prominent philosophical opponents.)

Beth's theorem and Craig's theorem

Beth's theorem is a central result about definability of non-logical symbols in classical first-order theories. It states that a symbol P is implicitly defined by a theory T if and only if an explicit definition of P in terms of some other expressions of the theory T can be deduced from the theory T . Intuitively, the symbol P is implicitly defined by T if, given the extension of these other symbols, T fixes the extension of the symbol P uniquely. In a precise statement of Beth's theorem this will be replaced by a condition on the models of T . An explicit definition of a predicate symbol states necessary and sufficient conditions: for example, if P is a one-place predicate symbol, an explicit definition is a sentence of the form $(x)(Px \equiv \phi(x))$, where $\phi(x)$ is a formula with free variable x in which P does not occur. Thus, Beth's theorem says something about the expressive power of first-order logic: there is a balance between the syntax (the deducibility of an explicit definition) and the semantics (across models of T the extension of P is uniquely determined by the extension of other symbols).

Beth's definability theorem follows immediately from Craig's interpolation theorem. For first-order logic with identity, Craig's theorem says that if ϕ is deducible from ψ , there is an interpolant θ , a sentence whose non-logical symbols are common to ϕ and ψ , such that θ is deducible from ψ , while ϕ is deducible from θ . Craig's theorem and Beth's theorem also hold for a number of non-classical logics, such as intuitionistic first-order logic and classical second-order logic, but fail for other logics, such as logics with expressions of infinite length.

1 The axiomatic method

Questions concerning the definability of concepts arose within the development of the formal axiomatic method by late nineteenth- and twentieth-century mathematicians such as Moritz Pasch, Giuseppe Peano, David Hilbert and Alfred Tarski. What sets apart the formal axiomatic method from the earlier axiomatic thinking of, say, Euclid's geometry is that the primitive terms occurring in the axioms are uninterpreted and the axioms devoid of any meaning; and that the rules of reasoning have been made completely explicit and formal. The axioms state purely formal relationships between the terms. This suffices for the purpose of deducing theorems from the axioms by rigorous reasoning. The development of the formal axiomatic method is the culmination of the movement to rigorize mathematics started in the early nineteenth century by mathematicians such as Cauchy and Bolzano. But these earlier efforts were directed chiefly towards ontology, for example, cleansing the language of analysis of visual images and reference to movement - and a mathematical theory still had subject matter, albeit abstract. In the formal axiomatic method the subject matter is provided from the outside by an interpretation of the primitive terms of the theory. One and the same theory may be open to radically diverse interpretations.

A perspicuous system of axioms requires under the formal axiomatic method the independence of each axiom from the others: a dependent axiom can be dropped without loss of content. Peano developed a method to prove independence: give an interpretation of the primitive terms which makes the one axiom false and the others true. This now common method is the formalization of an earlier idea of Eugenio Beltrami who felt that the so-called non-Euclidean geometry of Lobachevskii lacked a 'real foundation', that is a foundation in actual physical space. In 1868 Beltrami offered an interpretation of this geometry in terms of the acceptable Euclidean geometry: Lobachevskian 'geometry' could be understood as being about a special kind of line (a geodesic) in a special kind of plane (a surface with constant negative curvature) in Euclidean space. But Beltrami's project was one of meaning and he did not have the consistency of non-Euclidean geometry in mind nor was he concerned with the independence of Euclid's parallel axiom. In fact, only two years later, the French mathematician Guillaume Hoüel pointed out that Beltrami's construction showed the independence of the parallel axiom: while the other axioms are true for all planes, the parallel axiom holds only in planes of zero curvature (see Scanlan 1988).

Another idea from the pre-formal period was absorbed by the formal axiomatic method through the work of the Italian mathematician Alessandro Padoa, a close collaborator of Peano. Padoa was concerned with the definability of concepts. He pointed out that there exists a parallel between, on the one hand, the methodological concepts of being an axiom and being derivable as a theorem and, on the other hand, concepts from the theory of definition: the notion of primitive term corresponds to the notion of being an axiom, and the notion of definability corresponds to the notion of deducibility. A perspicuous system of axioms therefore also requires the

independence of the primitive concepts used in the axioms. Padoa was thus led to ask in 1900 whether there could be a method to prove independence of concepts, just as Peano's method had shown independence of axioms.

Early in the nineteenth century the French mathematician Jose Diez Gergonne had suggested the contrasting terms 'explicit' and 'implicit' as regards definition (Gergonne 1918-19). Gergonne's distinction was suggested by the difference, in algebra, between a set of solved equations, which gives as it were explicit definitions of the unknowns, and a set of unsolved equations which is strong enough to determine a unique solution for the unknowns. Gergonne characterizes implicit definitions as 'phrases that make us understand one of the words that occur in it through the known meaning of the other words'. Analogously to the algebraic case, Gergonne requires that the number of unknown words should be equal to the number of phrases that together implicitly define them. Completely forgotten by the end of the nineteenth century it was Giovanni Vacca, an assistant of Peano, who around 1896 gave a short account of Gergonne's paper on definitions. Padoa's paper (1901), read at the First International Congress of Philosophy in Paris, is clearly motivated by Gergonne's work. To prove the independence of a concept from the other concepts occurring in a theory Padoa proposed a new method: find a true interpretation of the theory, considered as an abstract system, that remains a true interpretation when solely the meaning of that concept is changed. Thus, though Padoa did not refer to Gergonne or use the term 'implicit definition', his so-called 'two-model method' establishes that a term is not implicitly defined by a theory in the sense of Gergonne.

Since Padoa did not indicate how to construct the two models it may be better to speak of the 'Padoa criterion' for undefinability. Padoa claimed without further proof that his criterion was both necessary and sufficient for the explicit undefinability of a given concept by means of the other concepts in a given theory. Sufficiency is clear: if an explicit definition were to be implied by the theory two such models could not exist since the truth of the explicit definition would force the uniqueness of the interpretation of the explicitly defined term given an interpretation of the other terms. But necessity is not obvious. Does the absence of two such models guarantee that an explicit definition exists and is derivable? With hindsight this question could not have been answered at that time for it requires a more careful specification of the underlying logical system than was available to Padoa.

Alfred Tarski (1935) answered the question affirmatively for a modification of the ramified theory of types of Whitehead and Russell's *Principia Mathematica* (see [Theory of types](#)). His proof was a rather straightforward derivation within the system, since the meta-claim that Padoa's two models do not exist can be expressed in the language of type theory.

2 Beth's theorem and Craig's interpolation theorem

In 1953 the Dutch philosopher and logician Evert Beth proved the necessity of Padoa's criterion for first-order or elementary logic. Beth showed that if no explicit definition of a term can be deduced from a theory, two models of the theory exist that differ only in the interpretation of the term in question. Moreover, in his so-called semantic tableau method, Beth found the means to construct systematically, in the absence of definability, the two models required by Padoa's criterion, albeit often through an infinite process, while a closed tableau makes it possible to find an explicit definition of the term in question (see [Natural deduction, tableau and sequent systems §4](#)). Beth thus took away some of the concerns of the American mathematician Oswald Veblen who had remarked in 1902 that what Padoa proposed 'seems hardly adequate' when the issue was to replace an axiomatic system by one with independent axioms and independent terms, since he gave no method to find the two models and thus prove independence, or to construct the explicit definition, in the case of dependency.

Let L be a first-order language and P an arbitrary non-logical constant not in L . Let $L(P)$ denote the language obtained by adding P to L . To simplify our notation we will assume that P is a one-place predicate symbol. If T is an arbitrary theory in the language L , then $T(P)$ will be a theory in the language $L(P)$. Deducibility in first-order logic will be denoted by ' $\vdash$ '. An interpretation or model M for L specifies extensions in a domain D for all the non-logical constants of L (see [Model theory](#)). If M is a model of L , a model of $L(P)$ will be denoted by (M, X) , where X is a subset of the domain D of M . Thus P is here interpreted as the subset X .

We will now define the notions of explicit and implicit definability. $T(P)$ is said to *define P explicitly* if there is a formula $\phi(x)$ of L such that $T(P) \vdash (x)(Px \equiv \phi(x))$. If P is not a one-place predicate symbol, this definition can be modified in the obvious way. Furthermore, let T be the set of first-order consequences of $T(P)$ in the

language L , in which P does not occur. Then $T(P)$ is said to *define P implicitly* when, for every model M of T , there is exactly one expansion (M, X) of M which is a model of $T(P)$.

'Beth's definability theorem' for first-order logic states that a theory $T(P)$ defines a term P implicitly if and only if $T(P)$ defines P explicitly. Beth's original proof of 1956 uses a modification of Gentzen's 'extended Hauptsatz', which shows that, in first-order logic, every proof can be carried out without any detours. Nowadays Beth's theorem is usually proved to be a direct implication of Craig's interpolation theorem.

'Craig's interpolation theorem' for first-order logic with identity says that if a sentence ψ of first-order logic entails a sentence θ there is an 'interpolant', a sentence ϕ in the vocabulary common to θ and ψ , that entails θ and is entailed by ψ . William Craig originally proved this theorem as a lemma to be used in obtaining a simpler proof of Beth's theorem (Craig 1957). Since then, however, the result has come to stand on its own.

We will now sketch a proof that Craig's theorem implies Beth's theorem. Since first-order logic is complete, implicit definability, a model-theoretic condition, is equivalent to the following deducibility condition, which is in fact Beth's original definition of (implicit) definability. Let P' be a one-place predicate not in L and distinct from P , and let $T(P')$ be the theory in $L(P')$ obtained by replacing P by P' in $T(P)$ wherever it occurs. Then P is implicitly defined by $T(P)$ if and only if $T(P) \cup T(P') \vdash (x)(Px \equiv P'x)$. Assume now that this condition holds, and that $T(P)$ is a finite set of sentences, or, rather, one big conjunction of axioms, and similarly for $T(P')$. So we can write $T(P) \& T(P') \vdash (x)(Px \equiv P'x)$. But then also $T(P) \& Pc \vdash (T(P') \rightarrow P'c)$, where c is a new individual constant not in L . By Craig's theorem there is an interpolant $\phi(c)$ such that $T(P) \& Pc \vdash \phi(c)$ and $\phi(c) \vdash (T(P') \rightarrow P'c)$. Since P' does not occur in $\phi(c)$ it is also true that $\phi(c) \vdash (T(P) \rightarrow Pc)$. Thus $T(P) \vdash (Pc \equiv \phi(c))$. Since c does not occur in $T(P)$, we also have $T(P) \vdash (x)(Px \equiv \phi(x))$. This completes the proof that if $T(P)$ defines P implicitly, then $T(P)$ defines P explicitly. The other direction of Beth's theorem follows independently of Craig's theorem.

3 Further developments and applications

Again, let T be the set of first-order consequences of $T(P)$ in the language L . We said that $T(P)$ defines P implicitly when for every model M of T there is exactly one expansion (M, X) which is a model of $T(P)$. In Padoa's method two models (M, X) and (M, X') of $T(P)$ are exhibited. Another way in which implicit definability can be violated is if there is a model of T that cannot in any way be expanded to a model of the full $T(P)$. Karel de Bouvere (1959) studied this so-called one-model method to show undefinability of addition and multiplication in number theory. In the philosophy of science literature this is called a failure of Ramsey eliminability of the term.

Not to be confused with the above concept of definability of a term in a theory is the concept of definability of a set in a model which, for Tarski, belongs to semantic definability rather than to the formal definability involved in Padoa's question since now we have a fixed model for an interpreted language. Given a model M of L , a subset X of its domain D is *definable in the model M* if there is a formula of L with one free variable $\phi(x)$ such that $(x)(Px \equiv \phi(x))$ is true in (M, X) , where P is interpreted as X . Obviously, if P is explicitly definable in $T(P)$ and if (M, X) is a model of $T(P)$, then X is definable in M . Moreover, the concept of definability in a model can be iterated, whereas definability of a term in a theory cannot since a set of terms is not itself a term of the language.

We say that the predicate P is definable in a model (M, X) of $T(P)$ if an explicit definition of P holds in (M, X) . Different models of $T(P)$ may satisfy different, non-equivalent definitions. But a theorem proven by Lars Svenonius in 1959 shows that if P is definable in every model (M, X) of $T(P)$ then each model (M, X) of $T(P)$ satisfies one of a finite list of definitions. That is, $T(P)$ implies a (finite) disjunction of explicit definitions of P . This property is called 'explicit definability up to disjunction' or 'piecewise definability'.

In model theory the concept of 'a logic' is defined and logics for which Craig's interpolation theorem holds are said to have the Craig or interpolation property; similarly for the Beth property. Any usual logic with the Craig property has the Beth property, but the latter has been shown to be weaker.

See also: [Definition](#); [Geometry, philosophical issues in](#); [Logical and mathematical terms, glossary of](#)

ZENO SWIJTINK

References and further reading

- Barwise, J. and Feferman, S.** (eds) (1985) *Model-Theoretic Logics*, New York: Springer. (The papers by Ebbinghaus and by Makowsky give details about the relation between the Beth property and the Craig property in abstract model theory.)
- Benthem, J.F. van** (1978) 'Ramsey Eliminability', *Studia Logica* 37: 321-36. (A detailed exposition of Ramsey eliminability with a discussion of its relevance for the philosophy of the empirical sciences.)
- Beth, E.W.** (1953) 'On Padoa's Method in the Theory of Definition', *Indagationes Mathematicae* 15: 330-9. (Gives Beth's result that for first-order logic implicit definability implies explicit definability in Beth's original proof-theoretic formulation.)
- Beth, E.W.** (1956) *L'Existence en Mathématiques*, Paris: Gauthier-Villars. (Includes an example of the one-model method to show undefinability. This involves constructing a model of the consequences of the theory without the terms whose undefinability has to be demonstrated, a model which cannot be extended to a model of the full theory.)
- Beth, E.W.** (1962) *Formal Methods: An Introduction to Symbolic Logic and to the Study of Effective Operations in Arithmetic and Logic*, Dordrecht: Reidel. (Shows how Beth's semantic tableau gives a method for constructing a Craig interpolant, and how this leads to the construction of an explicit definition in case Padoa's method fails.)
- Bouvére, K.L. de** (1959) *A Method in Proofs of Undefinability, With Applications to Functions in the Arithmetic of Natural Numbers*, Amsterdam: North Holland. (De Bouvére's dissertation with the one-model method, which shows that, under certain conditions, a term is essentially undefinable, that is, undefinable in any consistent extension.)
- Chang, C.C. and Keisler, H.J.** (1973) *Model Theory*, Amsterdam: North Holland. (Gives a complete account of interpolation and definition results for classical logic.)
- Craig, W.** (1957) 'Three Uses of the Herbrand-Gentzen Theorem in Relating Model Theory and Proof Theory', *Journal of Symbolic Logic* 22: 269-85. (Derives Craig's theorem about the existence of an 'intermediate' formula or interpolant from a consequence of Gentzen's extended *Hauptsatz*, and applies it in a proof of Beth's definability theorem, formulating it as a result relating model theory and proof theory.)
- Ebbinghaus, H.-D.** (ed.) (1987) *Bibliography of Mathematical Logic*, vol. 3, *Model Theory*, Berlin: Springer. (§C40 gives an extensive bibliography on interpolation and definability.)
- Gergonne, J.D.** (1918-19) 'Essai sur la théorie des définitions', *Annales de Mathématiques Pures et Appliquées* 9: 1-35. (Referred to in §1. Defines the concept of implicit definition before the advent of the formal axiomatic method.)
- Kleene, S.C.** (1967) *Mathematical Logic*, New York: Wiley. (A good account of the formal axiomatic method.)
- Padoa, A.** (1901) 'Essai d'une théorie algébrique des nombres entiers, précédé d'une introduction logique à une théorie déductive quelconque', in *Premier Congrès International de Philosophie*, vol. 3, *Logique et Histoire des Sciences*, Paris: Armand Colin, 309-65; partial trans. 'Logical Introduction to any Deductive Theory', in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967, 118-23. (Referred to in §1. The 'Logical Introduction on Deductive Theories in General' has the first statement of Padoa's criterion for the undefinability of a symbol on the basis of a theory.)
- Scanlan, M.J.** (1988) 'Beltrami's Model and the Independence of the Parallel Postulate', *History and Philosophy of Logic* 9: 13-34. (Referred to in §1. Compares the work of Beltrami and Hoüel with the treatment of Euclidean and non-Euclidean geometry after the development of the formal axiomatic method by Pasch, Peano and others.)
- Tarski, A.** (1935) 'Einige methodologische Untersuchungen über die Definierbarkeit der Begriffe', *Erkenntnis* 5: 80-100; trans. 'Some Methodological Investigations on the Definability of Concepts', in *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*, trans. and ed. J.H. Woodger, Oxford: Clarendon Press, 1956, 296-319. (Referred to in §1. Includes the proof that in the logical system of *Principia Mathematica* the two extensions required by Padoa's method do exist if a term cannot be explicitly defined.)
- Tarski, A. and Kuratowski, C.** (1931) 'Les opérations logiques et les ensembles projectifs' (Logical Operations and Projective Sets), *Fundamenta Mathematicae* 17: 240-8. (The first occurrence of the expression 'implicit definition' in the sense of Gergonne in the modern literature. The authors point out that definitions by induction are a type of implicit definition.)
- Tuomela, R.** (1973) *Theoretical Concepts*, Vienna: Springer. (An accessible discussion of definability and Ramsey eliminability in the empirical sciences.)

Vacca, G. (1896-9) 'Sui precursori della logica mathematica. II. J.D. Gergonne' (On Precursors of Mathematical Logic. II. J.D. Gergonne), *Rivista di matematica* 6: 183-6. (Referred to in §1. Brings the work of Gergonne on definition to the attention of the new logic.)

Boolean algebra

Boolean algebra, or the algebra of logic, was devised by the English mathematician George Boole (1815-64) and embodies the first successful application of algebraic methods to logic.

Boole seems to have had several interpretations for his system in mind. In his earlier work he thinks of each of the basic symbols of his ‘algebra’ as standing for the mental operation of selecting just the objects possessing some given attribute or included in some given class; later he conceives of these symbols as standing for the attributes or classes themselves. In each of these interpretations the basic symbols are conceived as being capable of combination under certain operations: ‘multiplication’, corresponding to conjunction of attributes or intersection of classes; ‘addition’, corresponding to (exclusive) disjunction or (disjoint) union; and ‘subtraction’, corresponding to ‘excepting’ or difference. He also recognizes that the algebraic laws he proposes are satisfied if the basic symbols are interpreted as taking just the number values 0 and 1.

Boole’s ideas have since undergone extensive development, and the resulting concept of Boolean algebra now plays a central role in mathematical logic, probability theory and computer design.

1 Basic facts

Boolean algebra, devised by the English mathematician George Boole, embodies the first successful application of algebraic methods to logic (see Boole 1847, 1854).

The algebraic structures implicit in Boole’s analysis were first explicitly presented by Huntington (1904), and termed ‘Boolean algebras’ by Sheffer (1913). As Huntington recognized, there are various equivalent ways of characterizing Boolean algebras. One of the most convenient definitions is as follows. A ‘Boolean algebra’ is a structure $(B, +_B, *_B, -_B, 0_B, 1_B)$, with B a non-empty set, $+$ and $*$ binary operations on B , $-$ a unary operation on B , and 0 , 1 distinct elements of B such that the following are true for all x, y, z in B .

$$x + (y + z) = (x + y) + z \quad (\text{associativity})$$

$$x * (y * z) = (x * y) * z$$

$$x + y = y + x$$

$$x * y = y * x$$

(commutativity)

$$x + (x * y) = x$$

(absorption)

$$x * (x + y) = x$$

$$x + (y * z) = (x + y) * (x + z)$$

(distributivity)

$$x * (y + z) = (x * y) + (x * z)$$

$$x + (-x) = 1$$

$$x * (-x) = 0$$

(complementation)

(We omit the subscript when confusion is unlikely.) The operations ‘Čplus;’ (Boolean ‘addition’ - here corresponding to *inclusive* disjunction), ‘Čast;’ (Boolean ‘multiplication’ - intersection) and ‘ $-$ ’ (taking the complement) are called Boolean operations. Note that the operation sending x, y to $x * (-y)$ corresponds to Boole’s ‘subtraction’. The elements 0 and 1 are called the ‘zero element’ and ‘unit element’ of B , respectively. We shall usually identify somewhat loosely a Boolean algebra by its underlying set; thus, for example, the Boolean algebra just introduced will be denoted simply by ‘ B ’.

The following basic identities can be readily shown to hold in any Boolean algebra.

$$-(x + y) = (-x) * (-y) \quad (\text{De Morgan’s laws})$$

$$-(x * y) = (-x) + (-y)$$

$$-(-x) = x \quad (\text{double complement law})$$

We observe that in the axioms characterizing Boolean algebras the roles played by $+$ and $*$, and by 0 and 1, are entirely symmetrical. This observation leads to the important ‘principle of duality’ for Boolean algebras which

may be formulated as follows. Let P be any statement about Boolean algebras which involves just the Boolean operations $+$, $*$ and $-$ and the elements 0 , 1 . The ‘dual’ of P is obtained from P by interchanging $+$ with $*$ and 0 with 1 . If P holds in all Boolean algebras, then so does the dual of P . Thus, for example, once we have established one of De Morgan’s laws above, the principle of duality leads automatically to the other one.

We define a ‘subalgebra’ of a Boolean algebra B to be a non-empty subset of B which is closed under the Boolean operations in B .

2 Important types of Boolean algebra

Boole seems to have had several interpretations for his algebraic system in mind. In his original pamphlet (1847) he thinks of each of the basic symbols of his ‘algebra’ as standing for the mental operation of selecting just the objects possessing some given attribute or included in some given class. Later (1854) he conceives of these symbols as standing for the attributes or classes themselves.

Two standard types of Boolean algebra arise in set theory and logic, corresponding to these two interpretations. To obtain the first of these, consider any non-empty set X and let $\wp(X)$ be its power set, that is, the set of all its subsets. Then the structure $(\wp(X), \cup, \cap, -, \emptyset, X)$ is a Boolean algebra - the power set algebra of X - in which $\cup, \cap, -$ are the operations of set-theoretic union, intersection and complementation with respect to X , respectively.

A subalgebra of a power set algebra is called an algebra of sets. As an example, for any set X , let $Z(X)$ be the set of all finite and cofinite subsets of X (a ‘cofinite’ set being the complement of a finite set). Then $(Z(X), \cup, \cap, -, \emptyset, X)$ is an algebra of sets called the finite-cofinite algebra of X .

When X is a singleton $\{a\}$, $\wp(X)$ reduces to the set $\{0, 1\}$, with $0 = \emptyset$ and $1 = X$. This algebra is called the ‘two-element Boolean algebra’ and is denoted by ‘2’. Its operations are displayed in the table below.

x	y	$x + y$	$x * y$	$-x$
0	0	0	0	1
0	1	1	0	1
1	0	1	0	0
1	1	1	1	0

The standard types of Boolean algebra arising in logic are the so-called Lindenbaum-Tarski algebras. To obtain these, we start with a theory T in a propositional or first-order language L . Define the equivalence relation $\approx$ on the set of formulas of L by $\phi \approx \psi$ if $T \vdash \phi \leftrightarrow \psi$. For each formula ϕ let $[\phi]$ be its $\approx$ -equivalence class. On the set $A(T)$ of such equivalence classes define the operations $+$, $*$, $-$ and the elements $1, 0$ by:

$$\begin{aligned} [\phi] + [\psi] &= [\phi \vee \psi] \\ [\phi] * [\psi] &= [\phi \wedge \psi] \\ -[\phi] &= [\neg \phi] \\ 1 &= [\alpha_0 \vee \neg \alpha_0] \\ 0 &= [\alpha_0 \wedge \neg \alpha_0], \end{aligned}$$

where α_0 is a fixed but arbitrary formula. Then the structure $(A(T), +, *, -, 0, 1)$ is a Boolean algebra, called the Lindenbaum-Tarski algebra of T . The Boolean algebra $A(\emptyset)$ (which depends solely on the language L) is denoted by $Alg(L)$ and is called the Lindenbaum-Tarski algebra of L .

Algebras of sets and Lindenbaum-Tarski algebras are typical Boolean algebras in the following sense. Call two Boolean algebras $(B, +_B, *_B, -_B, 0_B, 1_B)$ and $(C, +_C, *_C, -_C, 0_C, 1_C)$ ‘isomorphic’ if there is a one-one, onto map (that is, an ‘isomorphism’) $f: B \rightarrow C$ such that $f(0_B) = 0_C$, $f(1_B) = 1_C$ and, for any x, y in B ,

$$\begin{aligned} f(x +_B y) &= f(x) +_C f(y) \\ f(x *_B y) &= f(x) *_C f(y) \\ f(-_B x) &= -_C f(x). \end{aligned}$$

Then it can be shown that any Boolean algebra is isomorphic both to an algebra of sets and to the

Lindenbaum-Tarski algebra of some propositional theory. (Isomorphic Boolean algebras are structurally indistinguishable.) These facts (the first of which is the famous ‘Stone representation theorem’ of 1936) together show that Boolean algebras embody just the common structural features of set theory and logic.

Since, as is readily shown, there are Boolean algebras (for example, the finite-cofinite algebra of the set of natural numbers) which are *not* isomorphic to any power set algebra, nor to the Lindenbaum-Tarski algebra of a propositional language, it is of interest to try to determine the additional characteristic features of these two latter types of algebra.

For power set algebras, we need the following concepts. Given a Boolean algebra B , we define the relation $\leq$ on B by $x \leq y$ if and only if $x * y = x$. It is easily shown that $\leq$ is reflexive ($x \leq x$ for all x), transitive (if $x \leq y$ and $y \leq z$ then $x \leq z$) and antisymmetric (if $x \leq y$ and $y \leq x$ then $y = x$); in other words, $\leq$ is a ‘partial ordering’ on B . B is then said to be ‘complete’ if every subset X of B has both a greatest lower bound and a least upper bound with respect to $\leq$. (It is not hard to see that a *finite* subset $\{x_1, \dots, x_n\}$ of any Boolean algebra has greatest lower bound $x_1 * \dots * x_n$ and least upper bound $x_1 + \dots + x_n$.) An element $a \neq 0$ of a Boolean algebra B is called an ‘atom’ if, for any x in B , if $x \leq a$, then $x = 0$ or $x = a$. B is said to be ‘atomic’ if for any $x \neq 0$ in B there is an atom a such that $a \leq x$. Then we have the fundamental result of Lindenbaum and Tarski (1935) characterizing power set algebras: a Boolean algebra is isomorphic to a power set algebra if and only if it is complete and atomic (see Halmos 1963: §16, theorem 5).

Turning to Lindenbaum-Tarski algebras of propositional languages, we define a subset X of a Boolean algebra B to be ‘free’ if for any finite subset $\{x_1, \dots, x_n\}$ of X we have $y_1 * \dots * y_n \neq 0$, where each y_i is either x_i or $-x_i$. X is said to ‘generate’ B if the only subalgebra of B containing X is B itself, that is, if every element b of B can be expressed in the form $b = y_1 + \dots + y_n$, where each y_i is of the form $z_1 * \dots * z_m$ with either z_i or $-z_i$ in X . Finally X is said to be ‘freely generated’ if it has a free set of generators. It can then be shown that, for any propositional language L , the algebra $Alg(L)$ is freely generated (with the equivalence classes of the proposition letters constituting a free set of generators) and, moreover, that any (infinite) freely generated Boolean algebra is isomorphic to some $Alg(L)$.

3 Boolean algebras as algebras of truth-values

The simplest Boolean algebra is the algebra $2 = \{0, 1\}$, which it is customary to think of as the algebra of two ‘truth-values’; 0 standing for False and 1 for True. More generally, we may conceive of an arbitrary Boolean algebra as an algebra of ‘generalized truth-values’ - that is, containing elements corresponding to ‘truth-values’ different from True and False. In that case, it becomes natural to extend the usual idea of truth-valuations of (propositional) formulas (that is, in the algebra 2) to valuations in arbitrary Boolean algebras. Thus we define a ‘valuation’ of a propositional language L in a Boolean algebra B to be a map v from the set of formulas of L to B such that, for arbitrary formulas ϕ, ψ we have

$$v(\phi \vee \psi) = v(\phi) + v(\psi)$$

$$v(\phi \wedge \psi) = v(\phi) * v(\psi)$$

$$v(\neg\phi) = -v(\phi).$$

It can then be shown that, for any formula ϕ , the following assertions are equivalent:

- (1) ϕ is a propositional theorem.
- (2) $v(\phi) = 1$ for every valuation v in 2.
- (3) $v(\phi) = 1$ for any valuation v in any Boolean algebra.

(The equivalence between (1) and (2) is the completeness theorem for propositional logic.) A similar result may be established for predicate logic (extending the relevant completeness theorem), this time using the idea of a valuation in a complete Boolean algebra to enable quantified formulas to receive ‘truth-values’.

The idea of using Boolean algebras as algebras of truth-values scored what is regarded by logicians as its most spectacular success in the 1960s when it was shown by Robert Solovay and Dana Scott, and independently by Petr Vopenka, that Paul Cohen’s celebrated proofs of independence in set theory could be formulated in terms of what has come to be known as the method of ‘Boolean-valued models’ (see Bell 1985). Briefly, the method runs as

follows. Let L be the language of Zermelo-Fraenkel set theory (ZF; see [Set theory, different systems of](#)), and suppose that σ is a sentence of L whose formal independence of ZF we wish to establish - the continuum hypothesis, for example (see [Continuum hypothesis](#)). A complete Boolean algebra B is chosen and a class $V^{(B)}$ of sets - the Boolean universe induced by B - constructed. Let $L^{(B)}$ be the language obtained by adding to L a name for each element of $V^{(B)}$. Now one can construct a 'valuation' v of the (sentences of) the language $L^{(B)}$ in B : for each sentence ϕ of $L^{(B)}$, $v(\phi)$ is the element of B representing the 'Boolean truth-value' of ϕ in $V^{(B)}$. This map v is defined so as to map all the theorems of ZF to the unit element 1 of B : accordingly the 'structure' $(V^{(B)}, v)$ is called a Boolean-valued model of set theory. If the Boolean algebra B is chosen with finesse, it is possible to arrange things so that $v(\sigma) = 0_B$. That is, for a suitable choice of B , $V^{(B)}$ is a Boolean-valued model of set theory in which the given statement σ is false. This amounts to a proof of the independence of σ from the axioms of set theory. It follows that the problem of establishing the independence of a particular set-theoretic statement boils down in principle to demonstrating the existence of a complete Boolean algebra with specified properties. This method has been exploited with great success in recent years.

See also: [Boole, G.](#); [Logical and mathematical terms, glossary of](#)

J.L. BELL

References and further reading

- Bell, J.L.** (1985) *Boolean Valued Models and Independence Proofs in Set Theory*, Oxford Logic Guides 12, Oxford: Clarendon Press. (A detailed technical account of the subject.)
- Boole, G.** (1847) *The Mathematical Analysis of Logic, Being an Essay Towards a Calculus of Deductive Reasoning*, London: G. Bell & Sons; repr. in *Studies in Logic and Probability*, London: Watts, and LaSalle, IL: Open Court, 1952; also repr. in *An Investigation of the Laws of Thought*, New York: Dover, 1973. (The first appearance of Boole's system.)
- Boole, G.** (1854) *An Investigation of the Laws of Thought, on which are Founded the Mathematical Theories of Logic and Probabilities*, London: Walton & Maberley; repr. in *An Investigation of the Laws of Thought*, New York: Dover, 1973. (Further elaboration.)
- Hailperin, T.** (1976) *Boole's Logic and Probability*, Amsterdam: North Holland; 2nd edn, 1986. (An illuminating critical exposition of Boole's achievement from a modern standpoint.)
- Halmos, P.R.** (1963) *Lectures on Boolean Algebras*, Princeton, NJ: Van Nostrand. (An attractively written exposition of the basic theory of Boolean algebras.)
- Huntington, E.V.** (1904) 'Sets of Independent Postulates for the Algebra of Logic', *Transactions of the American Mathematical Society* 5: 288-309. (The first explicit postulate systems for Boolean algebra.)
- Monk, J.D.** (ed.) (1989) *Handbook of Boolean Algebras*, Amsterdam: Elsevier Science, 3 vols. (This set includes a series of articles on all mathematical aspects of Boolean algebras. Volume 1 is an excellent introduction to the subject.)
- Rasiowa, H. and Sikorski, R.** (1963) *The Mathematics of Metamathematics*, Warsaw: PWN. (A systematic technical account of the algebraic approach to logic.)
- Sheffer, H.M.** (1913) 'A Set of Five Independent Postulates for Boolean Algebras, With Application to Logical Constants', *Transactions of the American Mathematical Society* 14: 481-8. (Here it is shown that Boolean algebras can be axiomatized in terms of the single primitive operation corresponding to 'not both' - a fact apparently already known to Peirce.)

Cantor's theorem

Cantor's theorem states that the cardinal number ('size') of the set of subsets of any set is greater than the cardinal number of the set itself. So once the existence of one infinite set has been proved, sets of ever increasing infinite cardinality can be generated. The philosophical interest of this result lies (1) in the foundational role it played in Cantor's work, prior to the axiomatization of set theory, (2) in the similarity between its proof and arguments which lead to the set-theoretic paradoxes, and (3) in controversy between intuitionist and classical mathematicians concerning what exactly its proof proves.

Cantor's theorem states that the cardinal number ('size'; see [Set theory §3](#)) of the power set $\wp(A)$ of any set A (that is, the set of all subsets of A) is greater than the cardinal number of A itself.

One of the seemingly paradoxical properties of infinite sets is that they can be put into one-one correspondence with proper parts of themselves. For example, the function $f(n) = 2n$ maps each natural number to exactly one even number. If one-one correspondence is taken to indicate sameness of size for infinite sets, as it does for finite ones, then this means that the set of natural numbers must be said to be the same size as the set of even numbers, even though intuitively one would think that there should be twice as many natural numbers as there are even numbers. If any infinite set could be put into one-one correspondence with any other, there would be little point in trying to use 'numbers' as means of 'size' comparison for infinite sets. This is why Cantor's proof that there are infinite sets which cannot be put into one-one correspondence with each other was crucial to the development of his theory of transfinite ordinal and cardinal numbers.

Cantor was seeking a characterization of the structure of point continua and his original result (1874) was that no denumerable sequence of elements in an interval $[a, b]$ of the real line can contain all the elements of $[a, b]$. From this he concluded that a continuum of real numbers cannot be uniquely correlated with the natural numbers. To transform this result into what has become known as Cantor's theorem required making sense of the idea that every set, finite or infinite, has a cardinal number (sets between which there exists a one-one correspondence being deemed to have the same cardinal number); developing an arithmetic of infinite cardinal numbers; and establishing the existence of a one-one correspondence between the real numbers in any continuous interval and the set $\wp(\mathbb{N})$ of all subsets of the set $\mathbb{N}$ of natural numbers. The non-denumerability of the real numbers then becomes a special case of the more general set-theoretic result.

Russell was able to produce a contradiction in Frege's system of logic by considering the class of all classes that do not belong to themselves (see [Paradoxes of set and property §4](#)). A similar strategy is used to prove that there would be a contradiction if there were a one-one function f mapping a set A onto its power set $\wp(A)$. Consider $d = \{x \in A : x \notin f(x)\}$. This is a subset of A and hence $d \in \wp(A)$. Since f correlates every member of $\wp(A)$ with a unique member of A , there is a unique element $a_d \in A$ such that $d = f(a_d)$. But both $a_d \in d$ and $a_d \notin d$ entail contradictions. Since the existence of A and of $\wp(A)$ are not in question, there can be no such f which is both one-one and onto.

In the case where A is finite or denumerable, d coincides with the diagonal set, which is defined by interchanging 0's and 1's down the diagonal of a square array such as the following (in which 0 indicates that a_i belongs to $f(a_i)$ and 1 that it does not).

	$f(a_1)$	$f(a_2)$	$f(a_3)$	$f(a_4)$	...
a_1	0	1	1	0	...
a_2	0	0	1	1	...
a_3	1	1	1	0	...
a_4	1	0	0	1	...
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

While there is general agreement that such an argument can be used to establish, for example, that for any given enumeration of the real numbers it is possible to specify a real number which is not included in that enumeration, the step to concluding from this that there exists a set whose cardinal number is greater than that of the natural

numbers is disputed by those who, like the intuitionists, do not accept the existence of actually infinite totalities (see [Constructivism in mathematics](#); [Intuitionism](#)). Instead they may interpret the result as reflecting the difference between natural numbers and real numbers. The latter must be defined via potentially infinite sequences of rational numbers or potentially infinite decimal expansions, and so are never fully determinate objects. A function from a domain of such entities to a collection of determinate objects can never be guaranteed to be one-one. It is thus concluded that the real numbers form a potentially infinite domain of a different kind from that of the natural numbers, but not one which could be said to have a higher cardinality.

See also: [Continuum hypothesis](#); [Logical and mathematical terms, glossary of](#)

MARY TILES

References and further reading

- Cantor, G.** (1895, 1897) 'Beiträge zur Begründung der transfiniten Mengenlehre', parts 1 and 2, *Mathematische Annalen* 46: 481-512, 49: 207-46; trans. P.E.B. Jourdain (1915), *Contributions to the Founding of the Theory of Transfinite Numbers*, New York: Dover, 1955. (Two of Cantor's most important memoirs on transfinite numbers.)
- Dauben, J.W.** (1979) *Georg Cantor: His Mathematics and Philosophy of the Infinite*, Cambridge, MA: Harvard University Press. (An excellent intellectual biography which traces the development of Cantor's mathematical ideas, while putting them in the context of his philosophy of the infinite and his professional life.)
- Dummett, M.A.E.** (1978) *Truth and Other Enigmas*, London: Duckworth. (See the essays 'The Philosophical Significance of Gödel's Theorem', 'Platonism' and 'The Philosophical Basis of Intuitionistic Logic'.)
- Wilder, R.L.** (1952) *Introduction to the Foundations of Mathematics*, Krieger; 2nd edn, New York, London and Sydney: Wiley & Sons, 1965, 91-101. (Discussion of diagonal arguments.)
- Wittgenstein, L.J.J.** (1956) *Remarks on the Foundations of Mathematics*, ed. G.H. von Wright, R. Rhees and G.E.M. Anscombe, trans. G.E.M. Anscombe, Oxford: Blackwell, 2nd edn, 1967, part 1, appendix 2, 54-63; 3rd edn, 1978. (Includes a discussion of what exactly the proof of Cantor's theorem proves.)

Category theory, applications to the foundations of mathematics

Since the 1960s Lawvere has distinguished two senses of the foundations of mathematics. Logical foundations use formal axioms to organize the subject. The other sense aims to survey ‘what is universal in mathematics’. The ontology of mathematics is a third, related issue.

Moderately categorical foundations use sets as axiomatized by the elementary theory of the category of sets (ETCS) rather than Zermelo-Fraenkel set theory (ZF). This claims to make set theory conceptually more like the rest of mathematics than ZF is. And it suggests that sets are not ‘made of’ anything determinate; they only have determinate functional relations to one another. The ZF and ETCS axioms both support classical mathematics.

Other categories have also been offered as logical foundations. The ‘category of categories’ takes categories and functors as fundamental. The ‘free topos’ (see [Lambek and Couture 1991](#)) stresses provability. These and others are certainly formally adequate. The question is how far they illuminate the most universal aspects of current mathematics.

Radically categorical foundations say mathematics has no one starting point; each mathematical structure exists in its own right and can be described intrinsically. The most flexible way to do this to date is categorically. From this point of view various structures have their own logic. Sets have classical logic, or rather the topos Set has classical logic. But differential manifolds, for instance, fit neatly into a topos Spaces with nonclassical logic. This view urges a broader practice of mathematics than classical.

This article assumes knowledge of category theory on the level of [Category theory, introduction to §1](#).

1 Categorical thinking

The notion of ‘isomorphism’ aptly illustrates categorical thinking. In algebra two groups with the same group structure are ‘isomorphic’. Two spaces with the same topological structure are ‘homeomorphic’.

Before category theory, group isomorphism was defined group-theoretically, homeomorphism topologically, and so on. The definitions are unified in the categorical one (see [Category theory, introduction to](#) for an introduction to the concepts and notation). In any category, arrows $f : A \rightarrow B$ and $g : B \rightarrow A$ are called ‘inverse’ to one another if $gf = 1_A$ and $fg = 1_B$. Objects A and B are isomorphic if there are inverse arrows between them.

Category theory generally defines objects only up to isomorphism. That is, a categorical property will not have a unique instance in a category A but all its instances in A will be isomorphic and any object isomorphic to an instance will be an instance. For example, a ‘terminal object’ in a category A is an object T such that each object of A has exactly one arrow to T . A category A may have any number of terminal objects or none. Suppose T and T' are two such. Then there are arrows $T \rightarrow T'$ and $T' \rightarrow T$, and they are inverse since 1_T is the only arrow from T to T and similarly for T' . Hence T and T' are isomorphic. Conversely if T is terminal and there is an isomorphism $T \rightarrow T'$ then every object B of A has a unique arrow $B \rightarrow T \rightarrow T'$ to T' , so T' is terminal.

2 The elementary theory of the category of sets (ETCS)

The ‘elementary theory of the category of sets’ (ETCS) axiomatizes the category Set (see [Category theory, introduction to §1](#)). We give the axioms in the next section.

In ETCS a ‘singleton’ is any set terminal in Set. We pick one and call it 1 . We define an ‘element’ x of a set A to be a function from 1 to A , $x : 1 \rightarrow A$. Composing with any $f : A \rightarrow B$ gives an element of B , $fx : 1 \rightarrow B$. Trivially, each singleton has exactly one element.

We define a ‘product’ of sets B and C to be a set $B \times C$ and functions $p_1 : B \times C \rightarrow B$ and $p_2 : B \times C \rightarrow C$ such that for any set A and functions $f : A \rightarrow B$ and $g : A \rightarrow C$ there is a unique $u : A \rightarrow B \times C$ with $p_1 u = f$ and $p_2 u = g$. We can use $\langle f, g \rangle$ to name u , so $p_1 \langle f, g \rangle = f$ and $p_2 \langle f, g \rangle = g$. Thus elements $\langle x, y \rangle$ of $B \times C$ correspond to elements x of B and y of C .

Now 1 is only defined only up to isomorphism, as is the product $B \times C$, p_1 , p_2 . This is no laxity on our part. If a set S has any property expressible in ETCS without using names then so does every set isomorphic to S . Thus 1 is

distinguished from other singletons only by its name. We can pick a product of B and C to call $B \times C$, p_1 , p_2 but it is distinguished from other products of B and C only by its name.

The elements of an ETCS set are definite in number but without distinguishing properties. Each is merely an element of S and is itself and no other. Like Cantor's cardinals ETCS sets are distinguished only by cardinality and their elements are 'mere units' (see Lawvere 1994).

This is a point of controversy. In practice even in Zermelo-Fraenkel set theory (ZF) where sets can be uniquely defined we rarely bother to do so. We say 'Let S be a singleton' without saying which one. We speak of 'the product $B \times C$ ' without specifying its elements (that is, without choosing among the many ways to code ordered pairs $\langle b, c \rangle$ as sets in ZF). But some people claim this will not do in foundations (see Mayberry 1988).

Others favour such indeterminacy. Since Dedekind raised the issue for modern mathematics, some people claim mathematical objects should have only relevant properties. Natural numbers, say, should have only arithmetic properties. Every ZF set has non-arithmetic properties given by its particular elements. So on this view numbers cannot be ZF sets. But ETCS models of arithmetic have no irrelevant properties. So this view could say the numbers form an ETCS set (see Benacerraf 1965).

Specifically, 'recursion data' on a set S means an element x of S and a function $f : S \rightarrow S$. We define a 'natural number object' to be a set N with recursion data 0 and $s : N \rightarrow N$, called zero and successor, such that: for any recursion data x, f on any S there is a unique function $u : N \rightarrow S$ with

$$\begin{aligned} u(0) &= x \\ u(sn) &= f(u(n)) \text{ for all } n \in N \end{aligned}$$

Provably in ETCS there are infinitely many different natural number objects but all are isomorphic. None has any properties but the shared ones which follow from the definition (see McLarty 1994).

A 'subset' of an ETCS set A is any one-to-one function $i : S \rightarrow A$, that is, a function such that for all elements x and y of S , $ix = iy$ implies $x = y$. An element x of A is a 'member' of i if there is some element z of S with $iz = x$. We often write $S \rightarrow A$ for a subset, leaving the function i implicit. A 'relation' between sets A and B is a subset $R \rightarrow A \times B$ of the product. If $\langle x, y \rangle : 1 \rightarrow A \times B$ is a member of the subset we say x and y stand in the relation R or we write xRy .

A 'power set' of a set B is a set PB and an internal membership relation $\in^B \rightarrow B \times PB$ which ties subsets of B to elements of PB as follows: for each set A and relation $H \rightarrow B \times A$ there is exactly one function $h : A \rightarrow PB$, and vice versa, such that for all x in A , an element y of B has yHx if and only if $y \in^B hx$. Through H , each x in A defines a subset of B containing all y with yHx . Then hx is the element of PB tied to that subset. Power sets are defined up to isomorphism.

Now take any natural number object N and choose a power set PN , $\in^N$ for it, and then a power set for PN , and so on for as many levels as we want to work with. From these we can construct sets of rational, real and complex numbers in the usual way of arithmetized analysis. From here on ETCS works very much like ZF.

Mathematical practice, or even the typical set theory chapter in a maths text, relies on too few details and too much abuse of notation to say whether it is using ETCS or ZF. It can be cleaned up in either system. Yet ETCS claims to begin closer to practice. The basic ZF devices of transfinite accumulation, and coding pairs and numbers as sets, serve foundations only and are forgotten in practice. The basic devices of ETCS (the category axioms, finite limits, universal constructions - see Mac Lane 1986) are used throughout mathematics.

3 ETCS axioms and toposes

The topos axioms are the category axioms (see Category theory, introduction to §1) plus these two:

- (1) There is a terminal object 1 , and any objects A and B have a product $A \times B$, p_1 , p_2 .
- (2) Each object B has a power set PB , $\in^B$.

Products and power objects are defined as in §2, reading 'object/arrow' for 'set/function'. We do the same for natural number objects, and many toposes satisfy the further 'axiom of infinity':

(3) There is a natural number object $N, 0, s$.

The characteristic feature of **Set** is that each function is uniquely determined by its effect on elements:

(4) Given any f and g both $A \rightarrow B$, if for all $x : 1 \rightarrow A$ we have $fx = gx$, then $f = g$.

The ETCS axioms are the category axioms plus (1)-(4).

In general topos theory any arrow $f : B \rightarrow A$ may be called a ‘generalized element’ of A . We call an arrow $i : S \rightarrow A$ ‘monic’ if it is one-to-one for all generalized elements of S : for any object T and arrows $h : T \rightarrow S$ and $k : T \rightarrow S$, $ih = ik$ implies $h = k$. We call any monic $i : S \rightarrow A$ a ‘sub-object’ of A . The definition of subsets in §2 agrees with this in **Set** but not in general.

We say i is ‘included in’ another sub-object $j : P \rightarrow A$ if there is some $f : S \rightarrow P$ with $i = jf$. Trivially, $1_A : A \rightarrow A$ is monic and includes all sub-objects of A . Call it all of A . Every topos has an initial object ϕ , and for each A the unique $\phi \rightarrow A$ is monic and included in every sub-object of A . Call it empty. Any two sub-objects of A have an intersection, defined as the largest sub-object included in both. They have a union defined as the smallest including both. Each sub-object i has a negation (or pseudocomplement), defined as the largest sub-object whose intersection with i is empty.

The operations on sub-objects give each topos a natural logic. For details see Johnstone (1977: ch.5) or other introductions to topos theory. The law of noncontradiction holds in every topos since by definition the intersection of any sub-object with its negation is empty. In terms of logic, every formula $P \ \& \ \sim P$ is false. But in many toposes the union of a sub-object of A with its negation need not be all of A , and so a formula $P \vee \sim P$ need not be true. The law of the excluded middle (LEM) fails in some toposes.

For example, in the topos **Spaces** every object has a spatial structure and every arrow has a derivative. There is an object R which we think of as the line. The arrows $R \rightarrow R$ are the classical differentiable functions on the real numbers. The line has a point 0 , that is, an element $0 : 1 \rightarrow R$. So there is a sub-object $\{0\} \rightarrow R$ and its negation $\{x \neq 0\} \rightarrow R$. But the union $\{x|x = 0 \vee x \neq 0\}$ is not all of R . It is ‘all of R except the cohesion at 0 ’. An arrow from this union to R may lack a derivative at 0 and even be discontinuous at 0 simply because it can be defined separately on the pieces $\{0\}$ and $\{x \neq 0\}$. So LEM fails in **Spaces** for the formula $x = 0$, among many others.

Heyting’s first-order intuitionist logic is sound in **Spaces** as it is in every topos (see **Intuitionism §7**). It is not complete since in **Spaces** LEM holds for closed formulas. Every disjunction $P \vee \sim P$ with no free variables is true in **Spaces**. In **Set** the union of any subset and its negation is the whole set so LEM holds. Of course classical first-order logic is sound and complete in **Set**. Heyting’s logic is complete in some toposes, and so for logic in toposes collectively.

The axiom of choice is false in many toposes. It is easy to see it must fail in **Spaces**. Consider the function $f : R \rightarrow R$ defined by $fx = x^3 - x$. For each y in R there is an x with $fx = y$ so if the axiom of choice held there would be a choice function $g : R \rightarrow R$ such that for each y , $f(gy) = y$. But the graph of f shows no such function g can be continuous everywhere (let alone differentiable), so no such function can exist in **Spaces**. In fact the axiom of choice fails in any topos where LEM fails, since it implies LEM (see **Johnstone 1977** or other topos texts).

In another respect topos logic is not even formally intuitionistic. In the logic of any topos, for every object A , every formula with just one free variable of sort A defines a sub-object of A . Brouwer’s philosophical misgivings about set formation find no reflection here.

When Grothendieck created toposes for applications in topology, algebraic geometry and number theory (see **Category theory, introduction to §3**) he paid no attention to intuitionistic mathematics, let alone to Brouwer’s epistemology. And the theory today does not say reasoning must all be intuitionistic or constructive. Reasoning should be just as classical or constructive as will suit its context.

4 Radically categorical foundations

Lawvere offered ETCS as a logical foundation and to support an ontological claim. Since ETCS sets have no

determinate contents, ‘we seem to have partially demonstrated that even in foundations, not Substance but invariant Form is the carrier of the relevant mathematical information’. But ETCS is not his only logical foundation. For other purposes ‘a more satisfactory foundation will involve a theory of the category of categories’ (1964: 1,506, 1,510).

Lawvere (1966) axiomatized a ‘category of categories’. We call these axioms CCAF. They give general theorems of category theory without implying the existence of many specific categories. Lawvere gave various stronger axioms. Positing that Set exists is one. The axioms then give all the categories constructed from Set.

Another would be to posit a ‘category of all categories’ as one object in the CCAF universe which represents the whole universe. The set theory NF (see [Set theory, different systems of §7](#)) includes a set of all sets. It also has a category of all categories in the most naïve sense. But this set theory works so awkwardly with functions that the category of all categories lacks crucial properties (it is not Cartesian closed). Axiomatizing this does not seem promising. A more promising approach not yet much explored would use ‘fibered’ categories in the sense of Bénabou (1985).

The radically categorical position, though, does not lean on any one logical foundation. It says mathematical objects exist each in its own right. Given any type of structure one goal should be to describe it as intrinsically as possible, and the most intrinsic descriptions of structures to date are categorical. For example, axioms for the topos Spaces describe differential manifolds including the line R in their own terms. These axioms do not presume set theory or anything else. They are on a par with ETCS or ZF in giving a self-contained logical foundation for their subject. But axioms only organize our knowledge. Spaces, sets and more exist whether we axiomatize them or not.

If spaces are as real as sets in this sense, then the logic of Spaces is as genuine a logic as that of Set. This is the point of controversy over topos logic. The axioms for Spaces can be modelled in classical set theory. (Moerdijk and Reyes (1991) give many variants.) And Set can be modelled in Spaces (McLarty 1988). You cannot have one without the other. But you could claim the logic of sets is genuine logic while Spaces and other toposes give merely formal interpretations. The radically categorical position says otherwise: each topos (and other categories we do not go into here) has its own genuine logic.

Logical foundations can also describe relations between various branches of mathematics. It is useful in geometry to know how spaces relate to sets. Plus, modelling Spaces in Set and vice versa gives relative consistency results - the Spaces axioms are consistent if ETCS is and vice versa. Johnstone (1977: ch.9) gives relative consistency results for variants of ZF and ETCS. This can increase our confidence in all the axiom systems.

What no logical foundations can do, on this view, is give a starting point to justify the rest of mathematics. No such powerful axiom system is transparently consistent and there is no prospect of proving any one consistent (see [Gödel’s theorems](#)). The justification of mathematics rests on its whole place in our knowledge. All consistency proofs are relative consistency proofs.

See also: [Logical and mathematical terms, glossary of](#)

COLIN McLARTY

References and further reading

- Bénabou, J.** (1985) ‘Fibered Categories and the Foundations of Naïve Category Theory’, *Journal of Symbolic Logic* 50: 10-37. (Requires basic category theory. Cited in §4.)
- Benacerraf, P.** (1965) ‘What Numbers Could Not Be’, *Philosophical Review* 74: 47-73. (Discusses the classical problem for set-theoretic foundations that arithmetic has no unique set-theoretic representation. Requires basic logic.)
- Corry, L.** (1992) ‘Nicolas Bourbaki and the Concept of Mathematical Structure’, *Synthese* 92: 315-48. (Includes Bourbaki’s internal debates over category theory.)
- Johnstone, P.** (1977) *Topos Theory*, London: Academic Press. (The standard reference. Chapters 1-4 are quite demanding. Others are readable with basic topos theory. Cited in §§3-4.)
- Lambek, J. and Couture, J.** (1991) ‘Philosophical Reflections on the Foundations of Mathematics’, *Erkenntnis* 34: 187-209. (Claims moderate logicism, formalism, Platonism and intuitionism are all realized in the ‘free topos’. Requires basic formal logic.)
- Lawvere, F.W.** (1964) ‘An Elementary Theory of the Category of Sets’, *Proceedings of the National Academy of Sciences* 51: 413-418.

Science 52: 1,506-11.(Requires basic category theory.)

Lawvere, F.W. (1966) 'The Category of Categories as a Foundation for Mathematics', in S. Eilenberg (ed.) *Proceedings of a Conference on Categorical Algebra in La Jolla*, Berlin: Springer, 1-21.(Cited in §4. Requires basic category theory.)

Lawvere, F.W. (1969) 'Adjointness in Foundations', *Dialectica* 23: 281-96.(Includes a rapid introduction to the terms.)

Lawvere, F.W. (1994) 'Cohesive Toposes and Cantor's "lauter Einsen"', *Philosophia Mathematica* 2 (3): 5-15. (Describes aspects of Cantor's thought, such as his cardinals, lost in ZF set theory. Cited in §2.)

Mac Lane, S. (1986) *Mathematics Form and Function*, New York: Springer, 386-402.(Easy account of the categorical terms in §2.)

Marquis, J.-P. (1995) 'Category Theory and the Foundations of Mathematics: Philosophical Excavations', *Synthese* 103: 421-47.(Refers to much mathematics but readable without relevant background.)

Mayberry, J. (1988) 'What are Numbers?', *Philosophical Studies* 54: 317-54.(Argues that working methods in mathematics since Newton have disguised the role of concrete foundations, as in §2.)

McLarty, C. (1988) 'Defining Sets as Sets of Points of Spaces', *Journal of Philosophical Logic* 17: 75-90. (Compares the nineteenth-century creation of set theory from analysis to the relation between Set and Spaces. Requires basic logic. Cited in §4.)

McLarty, C. (1994) 'Numbers Can Be Just What They Have To', *Noûs* 27: 487-98.(Shows the structuralism of ETCS as in §2. No technical knowledge is required.)

Moerdijk, I. and Reyes, G. (1991) *Models for Smooth Infinitesimal Analysis*, New York: Springer. (Many variants on the Spaces of §3-4. Requires some differential geometry plus basic topos theory or skill in algebra.)

Category theory, introduction to

A ‘category’, in the mathematical sense, is a universe of structures and transformations. Category theory treats such a universe simply in terms of the network of transformations. For example, categorical set theory deals with the universe of sets and functions without saying what is in any set, or what any function ‘does to’ anything in its domain; it only talks about the patterns of functions that occur between sets.

This stress on patterns of functions originally served to clarify certain working techniques in topology. Grothendieck extended those techniques to number theory, in part by defining a kind of category which could itself represent a space. He called such a category a ‘topos’. It turned out that a topos could also be seen as a category rich enough to do all the usual constructions of set-theoretic mathematics, but that may get very different results from standard set theory.

1 Examples and the category axioms

Consider sets and functions. Write $f : A \rightarrow B$ to say f is a function defined on the set A and all its values are in the set B . Each set A has an identity function $1_A : A \rightarrow A$ with $1_A(x) = x$ for every x in A . Given any $f : A \rightarrow B$ and $g : B \rightarrow C$ there is a composite $gf : A \rightarrow C$ with $gf(x) = g(f(x))$ for all x in A .

Now consider a formal language with deductive rules. Let $f : A \rightarrow B$ say f is a proof whose sole assumption is the formula A and conclusion is the formula B . Each formula A gives a one-line proof $1_A : A \rightarrow A$ with A as assumption and conclusion. Proofs $f : A \rightarrow B$ and $g : B \rightarrow C$ concatenate to give $gf : A \rightarrow C$. We view a proof as transforming its assumption into its conclusion.

The features which are common to these two examples are widespread in mathematics. Category theory gives an axiomatic treatment of them. The category axioms speak abstractly of ‘objects’ and ‘arrows’: each arrow has an object as domain and another as codomain. We write $f : A \rightarrow B$ to say f is an arrow with domain A and codomain B . If the codomain of one arrow is the domain of another, as in the case of $f : A \rightarrow B$ and $g : B \rightarrow C$, then the arrows have a ‘composite’ $gf : A \rightarrow C$. Each object A has an ‘identity arrow’ $1_A : A \rightarrow A$. The axioms require that for any arrows $f : A \rightarrow B$, $g : B \rightarrow C$ and $h : C \rightarrow D$ we have $h(gf) = (hg)f$, $f1_A = f$ and $1_B f = f$.

Thus we have the ‘category’ Set, whose objects are sets and whose arrows are functions, and the category Proof, whose objects are formulas and whose arrows are proofs (see Lambek 1994). The axioms clearly hold for Set and for Proof. Other useful categories include: Top, with topological spaces for objects and continuous functions for arrows; and Group, with groups (as in abstract algebra) for objects and group homomorphisms for arrows.

Picture a category as a network of arrows between points. So Set is an infinite network with a point for each set and an arrow for each function. Each path of arrows connected head-to-tail has a composite going from the first point on the path to the last.

2 Working with categories

We can look at some simple categorical structures. An object T of a category A is a ‘terminal object’ in A if each object of A has exactly one arrow to T . An object W of A is a ‘weak terminal object’ in A if each object of A has at least one arrow to W .

Terminal objects of Set are singleton sets. Weak terminal objects of Set are non-empty sets. In Proof the weak terminal objects are the provable formulas. If W is provable with no assumptions then it is provable from any assumption. Nothing is provable in only one way, so Proof has no terminal object.

Reversing arrows in the definition of a terminal object defines an ‘initial object’: an object I of A such that each object of A has exactly one arrow from I . Similarly for a ‘weak initial object’. A set is initial in Set if and only if it is empty, and is weakly initial on the same condition. No formula is initial in Proof. Inconsistent formulas are weakly initial.

These definitions deal only with patterns of arrows. Yet in Set and Proof they coincide with specific concepts such as non-emptiness or provability. Other definitions give richer structures such as Cartesian products or power sets

in Set, the several connectives in Proof, and topological or group structures in Top or Group. So category theory has a unifying role where one argument proves different theorems in different contexts. More importantly, it focuses on the relevant structural relations between objects, be they sets or spaces or whatever. To see this at work we ascend to transformations of categories, called ‘functors’.

We write $F : \mathbf{A} \rightarrow \mathbf{B}$ to say F is a functor from the category $\mathbf{A}$ to the category $\mathbf{B}$. This means that for each object A of $\mathbf{A}$ there is an object FA of $\mathbf{B}$, and for each arrow $f : A \rightarrow A'$ of $\mathbf{A}$ there is an arrow $Ff : FA \rightarrow FA'$ of $\mathbf{B}$, and F preserves identity arrows and composites. That is, for each A of $\mathbf{A}$ we have $F(1_A) = 1_{FA}$, and for any composite gf we have $F(gf) = (Fg)(Ff)$. So F draws a picture of the network A in the network $\mathbf{B}$.

Each category $\mathbf{A}$ has an identity functor $1_{\mathbf{A}} : \mathbf{A} \rightarrow \mathbf{A}$ with $1_{\mathbf{A}}(A) = A$ and $1_{\mathbf{A}}(f) = f$ for all objects A and arrows f of $\mathbf{A}$. Functors compose in the obvious way. So categories and functors themselves are the objects and arrows of a category of categories (see [Category theory, applications to the foundations of mathematics §4](#)).

Functors were invented to focus on relevant structure. Since about 1900 homology theory has been central to topology. To simplify somewhat, by 1940 homology theory was a way of taking each topological space S and constructing from it a family of groups $H_n(S)$, one for each natural number n . By a further construction each continuous function $f : S \rightarrow S'$ between topological spaces gives homomorphisms $H_n(f) : H_n(S) \rightarrow H_n(S')$. So group theory could be used to prove theorems of topology. But there were many different ways to construct the groups. The subject was thriving and chaotic.

Eilenberg and Mac Lane looked away from constructing the groups $H_n(S)$ and saw homology instead as drawing the network of continuous functions in the network of homomorphisms. They saw each H_n as a functor $H_n : \mathbf{Top} \rightarrow \mathbf{Group}$. To do this they created category theory and wrote their classic paper, ‘[The General Theory of Natural Equivalences](#)’ (1945). Category theory is now the standard language for homology. It cleared away old complexities and made room for vast new ones.

The new homology (and related cohomology) spread to algebra and number theory taking category theory with it. Then category theory took other new directions. (Applications before 1970 are surveyed in passing in Mac Lane (1971).)

3 Toposes

Around 1960 Grothendieck attacked some famous problems in geometry and number theory. His project achieved its initial aims in 1973 with Deligne’s proof of the Weil conjectures, certain theorems concerning integer roots of polynomials (see [McLarty 1990](#)). For this project Grothendieck created ‘toposes’ (or, ‘topoi’). These were categories which he viewed as spaces. Every ordinary topological space can be represented by a topos but toposes are considerably more general than that. The topological method of cohomology can be applied to all toposes and thus Grothendieck defined the ‘étale’ cohomology that Deligne used. This cohomology is now central to number theory but not all its applications require toposes.

About 1970 Lawvere and Tierney gave first-order axioms which apply to all of Grothendieck’s toposes plus some other categories. Today we use ‘topos’ or ‘elementary topos’ to mean any model of the axioms, and ‘Grothendieck topos’ to mean one got by Grothendieck’s method. To a logician a topos is analogous to the universe of sets. In any topos the constructions of ordinary mathematics exist, but not always in classical form. Classical logic is not valid in all toposes.

The category Set is a topos. Classical logic is valid in it and it gives classical mathematics. But there is also a topos Eff in which all functions from the natural numbers to themselves are recursive. Classical logic would prove there are non-recursive functions; but classical logic is not valid in Eff. There is a topos Spaces in which each object has a spatial structure and each function is differentiable. This topos includes all the usual differentiable functions of calculus. Classical logic proves there are non-differentiable functions but it is not valid in Spaces. In any topos the valid logic is somewhere between classical and intuitionistic. This is uncontroversial as a fact in model theory. For controversy over its philosophical interpretation see [Category theory, applications to the foundations of mathematics §4](#).

See also: [Logical and mathematical terms, glossary of](#)

COLIN McLARTY

References and further reading

- Bell, J.L.** (1988) *Toposes and Local Set Theories*, Oxford: Oxford University Press.(A logician's account of toposes using their analogy with the classical universe of sets. Requires facility with sets and formal logic. Includes a bibliography.)
- Eilenberg, S. and Mac Lane, S.** (1945) 'The General Theory of Natural Equivalences', *Transactions of the American Mathematical Society* 58: 231-94.(The original paper. Requires basic knowledge of vector spaces and topology.)
- Lambek, J.** (1994) 'What is a Deductive System?', in D.M. Gabbay (ed.) *What is a Logical System?*, Oxford: Oxford University Press, 145-64.(Surveys styles of proof theory including the categorical one in §1.)
- Mac Lane, S.** (1971) *Categories for the Working Mathematician*, Berlin: Springer.(The standard. The main text requires some sophistication; the examples call for much background. Includes a thorough bibliography.)
- Mac Lane, S.** (1986) *Mathematics Form and Function*, New York: Springer, 386-402.(Basic category theory, including Set, is described with no substantial prerequisites.)
- Mac Lane, S. and Moerdijk, I.** (1992) *Sheaves in Geometry and Logic*, Berlin: Springer.(The general theory of Grothendieck toposes, with many examples, including Cohen's forcing as a topos construction. Requires some category theory.)
- McLarty, C.** (1990) 'The Uses and Abuses of the History of Topos Theory', *British Journal for the Philosophy of Science* 41: 351-75.(Debates some views of the history of toposes.)
- McLarty, C.** (1992) *Elementary Categories, Elementary Toposes*, Oxford: Oxford University Press.(A categorical account, with chapters on Set, Spaces and Eff. Requires facility with algebraic thinking and formal logic. Includes a bibliography.)

Chaos theory

Chaos theory is the name given to the scientific investigation of mathematically simple systems that exhibit complex and unpredictable behaviour. Since the 1970s these systems have been used to model experimental situations ranging from the early stages of fluid turbulence to the fluctuations of brain wave activity. This complex behaviour does not arise as a result of the interaction of numerous sub-systems or from intrinsically probabilistic equations. Instead, chaotic behaviour involves the rapid growth of any inaccuracy. The slightest vagueness in specifying the initial state of such a system makes long-term predictions impossible, yielding behaviour that is effectively random. The existence of such behaviour raises questions about the extent to which predictability and determinism apply in the physical world. Chaos theory addresses the questions of how such behaviour arises and how it changes as the system is modified. Its new analytical techniques invite a reconsideration of scientific methodology.

1 Defining chaos theory

Although there is controversy over the proper scientific definition of chaotic behaviour, chaos theory may be defined as the qualitative study of unstable aperiodic behaviour in deterministic nonlinear dynamical systems.

A dynamical system is a mathematical entity that usually results from the attempt to model a situation in the natural world (see **Models**). It has two components: a procedure for producing a mathematical description of the instantaneous state of a physical system, and a rule (the evolution equations) for transforming the current state description into a description for some other time. For some dynamical systems, the evolution equations yield a simple formula that allows the direct calculation of future states of the system.

The dynamical systems of interest for chaos theory contain nonlinear terms in the equations. These terms, such as x^2 or $\sin(x)$ or $5xy$, usually render an exact solution impossible. Researchers into chaotic phenomena instead seek a qualitative account of the system's behaviour, investigating the general character of its long-term behaviour, rather than seeking to arrive at numerical predictions of its exact future state (this special sense in which chaos theory is 'qualitative' must not be understood as contrasting with 'mathematical'). For example, a qualitative study of a planetary system would seek to discover what circumstances will lead to elliptical orbits as opposed to hyperbolic ones. This approach is called 'dynamical systems theory' (or sometimes simply 'dynamics') and began with the work of Henri **Poincaré**. Chaos theory is a specialized application of dynamical systems theory that typically treats such questions as: what characteristics will all solutions of this system ultimately exhibit? and, how does this system change from exhibiting one kind of behaviour to another kind?

Dynamical systems are usually studied in terms of trajectories in state space (sometimes called 'phase space'), a mathematically constructed abstract space where each dimension corresponds to one variable of the system. Every point in state space represents a full description of the system, and the evolution of the system manifests itself as the tracing out of a path, or trajectory, in state space. It is often possible to study the geometric features of these trajectories, even in actual experimental systems, without explicit knowledge of the solutions they represent. This allows one to characterize trajectories according to their topological features, and the investigation of how those features change as the parameters of the dynamical system are altered.

While qualitative questions can be asked about almost any dynamical system, chaos theory focuses on certain forms of behaviour - behaviour which is unstable and aperiodic. The form of instability known as sensitive dependence on initial conditions is a distinguishing characteristic of chaotic behaviour. A dynamical system that exhibits sensitive dependence on initial conditions will produce markedly different trajectories for two initial states that are initially very close together. In fact, given any specification of initial conditions, there is another set of initial conditions arbitrarily close to it that will diverge from it by some finite distance, given enough time. In most chaotic behaviour, sensitive dependence on initial conditions results from the exponential growth of any small initial difference. It is this instability that makes chaotic behaviour unpredictable (see §2).

Aperiodic behaviour occurs when no variable describing a property of the system undergoes a regular repetition of values. Furthermore, chaos is an appropriate label only when such behaviour occurs in a bounded system. An explosion thus does not qualify as chaotic behaviour. And systems studied by chaos theory bear the label

‘deterministic’ because their equations make no explicit reference to chance mechanisms (see §3).

For a dissipative dynamical system, characterized by the gradual loss of energy, trajectories in state space will asymptotically approach a shape known as an ‘attractor’ when transient effects have died away. Until the advent of chaos theory, only three types of attractor were generally recognized: the fixed point (corresponding to eventual equilibrium), the limit cycle (corresponding to periodic behaviour), and the torus (corresponding to behaviour with multiple periods in rationally incommensurable ratios). Chaotic behaviour in dissipative systems requires the introduction of a fourth variety: the so-called ‘strange attractor’.

A crucial geometric feature of strange attractors is their combination of stretching and folding. The action of such a system takes nearby points and ‘stretches’ them apart in a certain direction, thus creating the exponential divergence responsible for unpredictability. But the system also acts to ‘fold together’ points that are at some distance, bringing about convergence of trajectories in a different direction, and hence asymptotic attraction.

The stretching and folding of chaotic systems gives strange attractors the distinguishing characteristic of a nonintegral, or fractal, dimension. These attractors often appear as stacks of two-dimensional sheets packed in a self-similar structure that seems to intrude into three-dimensional space. The dimension of such an object - more than two but less than three - describes its scaling properties, giving a quantitative indication of the stretching and folding at work in the dynamical system. Another quantitative characterization of chaotic systems is given by the Lyapunov exponents, which measure the degree of sensitivity to initial conditions and thus the degree of unpredictability.

It should be noted that strange attractors appear only in dissipative systems. For systems where energy is conserved, there is no convergence onto an attractor. Instead, trajectories are confined to a surface of constant energy in state space. Chaotic behaviour can occur in such systems, but the trajectories will fill the allowed energy surface (which may itself display an extremely complicated structure).

The examination of the transition to chaotic behaviour has yielded important results for understanding the way complex behaviour arises. For example, period-doubling behaviour is found in a wide variety of nonlinear systems, and this type of transition has certain universal features. This universality allows us to make exact predictions about a system’s period-doubling behaviour as soon as we have verified that certain qualitative conditions hold.

Sensitive dependence on initial conditions, strange attractors and other elements of chaos theory have been reported in experimental systems in the physical and biological sciences. Some of these examples of chaotic behaviour have been convincingly documented in laboratory settings, while some of the examples of chaos outside the laboratory are the subject of lively debate.

2 Predictability

Moderately accurate initial conditions will always allow us to predict the chaotic behaviour of a system for some short time into the future, but since the vagueness in our initial conditions grows exponentially with time, we can see that for some chaotic systems, a useful prediction would require initial conditions to be specified with more accuracy than is possible. Inaccuracy can be due to equipment error, ineliminable perturbations from outside influences, round-off error in the representation of real numbers, or even the inherent indeterminacy of the quantity being measured.

Some may contend that so long as the prediction task for a chaotic system requires only some finite degree of accuracy, it is incorrect to say that such accuracy is impossible to achieve.

Since chaos theory, unlike special relativity or quantum mechanics, introduces no new fundamental postulates about the physical world, it provides no new physical theory that rules out the possibility of such predictions. Yet these same systems may nevertheless be unpredictable *in principle* because our inability to make such calculations in the longer term is not due to some constraint we could overcome. The very finitude of human resources implies that it is impossible to obtain the accuracy necessary for predictions.

The limitation posed by chaotic systems presents a challenge to the methodological assumption that a small amount of vagueness in measurements will lead to only a small amount of vagueness in predictions. The standard

technique of perturbation analysis treats a difficult case as a version of a predictable system plus some small ‘perturbation’ that complicates the situation but keeps it mathematically tractable. Sensitive dependence guarantees that such approximative techniques will be worthless, thus calling into question some of the metaphysical beliefs that underwrite their general applicability. One such belief is the conviction that systems governed by simple rules must behave simply.

3 Determinism

Scientists and philosophers use the term ‘determinism’ in several senses (see [Determinism and indeterminism §1](#)). By one criterion, a system is deterministic if the way it changes in time can be specified by a set of equations that make no reference to chance. That is, there are no probabilities built into the state description due to an averaging over many indistinguishable substates, and the evolution equations do not involve probabilistic branching. Thus, the systems treated by chaos theory are deterministic by definition.

Another criterion consists in the uniqueness of a system’s evolution: if two systems are identical in physical properties at one time, they must remain identical at all other times. That is, the complete instantaneous description of a deterministic system ‘fixes’ the past and future with no alternatives.

A third notion of determinism is that a system is predictable in principle by an all-powerful intelligence or computational scheme, given complete information of instantaneous conditions and the complete set of physical laws. Since measurements and calculations cannot be made with unlimited accuracy, a chaotic system will be unpredictable even though it satisfies other requirements for determinism. In other words, we may not be able to tell *which* unique trajectory a system is following, but that does not mean such trajectories do not exist.

On one interpretation, the randomness and apparent indeterminism of chaotic behaviour are merely artifacts of our epistemic limitations. Another interpretation would treat unique trajectories as inadmissible idealizations, arguing against uniqueness of evolution in chaotic behaviour. This instrumentalist approach argues that there is no point in insisting that the future is fixed, given the present, if that future cannot be predicted.

A third approach would draw on the implications of quantum mechanics to show that unique evolution encounters difficulty due to the workings of chaotic dynamics on intrinsically indeterminate quantities. Such an approach would require an account of how to combine the Newtonian formulations of nonlinear dynamics with the mathematical formalism of quantum mechanics - the difficult task of making sense of ‘quantum chaos’.

4 Chaos and scientific understanding

One of the apparent paradoxes of chaos theory is that it represents a scientific study of unpredictable systems that actually has significant predictive power. Chaos theory provides an understanding of how predictable large-scale or long-term patterns appear in behaviour which is nonetheless unpredictable in detail. It may seem odd to assert that chaos theory has ‘great predictive power’, but such claims refer not to an ability to predict the exact value of some property of a system, but to an ability to foresee and understand changes in the overall behaviour of that system. The crucial point here is the distinction between specific *quantitative* predictions, which are typically impossible for chaotic systems, and *qualitative* predictions, which are at the heart of dynamical systems theory.

The fruitfulness of chaos theory militates against the methodological assumption that it is always appropriate to seek to understand the behaviour of a system by trying to determine the equations governing the interactions of its parts (see [Reduction, problems of](#)). For most nonlinear dynamical systems, such a reductionist strategy does not help us understand how the chaotic behaviour sets in, or what kind of attractor characterizes it, or how the system will respond to changes in parameters. Understanding this behaviour requires the use of the new mathematical techniques of nonlinear dynamics. While chaos theory argues against the universal applicability of the method of reductionism, it provides no examples of ‘holistic’ properties which could serve as counter examples to metaphysical doctrines of reductionism, such as physicalism.

Chaos theory shows us the mechanisms responsible for unpredictable behaviour, but these mechanisms are best characterized as geometrical rather than causal. Systems with very different underlying causal substrata are grouped together in order to study their qualitative behaviour. Chaos theory does not need to reveal the causal process at work, and it would be impossible in principle to trace out the workings of the actual causal mechanism

in a chaotic system. To do this would require calculations using the evolution equations of the system.

According to some accounts, science furnishes understanding by widening the scope of events that happen out of necessity and not contingency - by revealing laws of nature and subsuming phenomena under them (see [Explanation §2](#)). While chaos theory establishes powerful generalizations, its practitioners do not portray their work as discovering new laws of nature. For instance, there is no set of necessary and sufficient conditions that would allow us to foresee which type of transition to chaotic behaviour a system will follow. Chaos theory emphasizes finding patterns and connections, while jettisoning the requirement that the patterns must yield necessity in a detailed and deterministic sense. Some may contend that this search for patterns actually constitutes striving to discover new laws governing qualitative features of systems.

Perhaps chaos theory does not provide us with any genuine understanding at all. After all, it does not introduce any fundamental revisions in our laws of nature. Some might argue that it merely introduces stop-gap measures for data analysis in systems with intractable equations. Perhaps it is only a source of new observations or empirical generalizations. But nonlinear dynamics does require a wholesale revision of notions about the applicability of classical models, and provides powerful new analytic techniques. It should not be disparaged on the grounds that fundamental theoretical structures are left unchanged.

5 Applications outside the natural sciences

Chaos theory has attracted attention from outside the natural sciences. In the social sciences, the techniques developed for analysis of nonlinear systems have been applied to economic data, for instance. Such techniques can determine the extent to which random fluctuations in a chosen variable can be modelled with simple deterministic mechanisms.

Sensitive dependence on initial conditions has also served as a metaphor for the precarious swings of human history. Metaphorical interpretations of chaos theory have also appeared in natural theology and literary theory. Other areas where chaos theory may have philosophical implications are: the foundations of statistical mechanics, the problem of freedom of the will, and the role of the cultural context of science.

See also: [Randomness](#)

STEPHEN H. KELLERT

References and further reading

- Batterman, R.W.** (1991) 'Chaos, Quantization, and the Correspondence Principle', *Synthese* 89: 189-227. (A useful discussion of some of the issues surrounding 'quantum chaos'.)
- Bergé, P., Pomeau, Y. and Vidal, C.** (1984) *Order Within Chaos*, trans. L. Tuckerman, Paris: Wiley & Sons. (Useful guide to chaos in the physical sciences.)
- Devaney, R.** (1986) *An Introduction to Chaotic Dynamical Systems*, Menlo Park, CA: Benjamin-Cummings Publishing. (A standard presentation of the mathematics of chaos at the introductory level.)
- Dyke, C.** (1990) 'Strange Attraction, Curious Liason: Clio Meets Chaos', *Philosophical Forum* 21: 369-92. (Examines the implications of chaos theory for history and the social sciences.)
- Earman, J.** (1986) *A Primer on Determinism*, Dordrecht: Reidel. (A comprehensive examination of philosophical issues about scientific determinism. Includes a technical discussion of chaos.)
- Ford, J.** (1989) 'What is Chaos, That We Should be Mindful of It?', in P. Davies (ed.) *The New Physics*, Cambridge: Cambridge University Press, 348-72. (Presents a number of arguments and speculations regarding predictability and determinism as they relate to chaos theory.)
- Hao, B.-L.** (ed.) (1984) *Chaos*, Singapore: World Publishing Company. (A useful collection of historically important scientific papers on chaos.)
- Hobbs, J.** (1991) 'Chaos and Indeterminism', *Canadian Journal of Philosophy* 21: 141-64. (Presents an argument on the relevance of chaos theory to issues of determinism.)
- Kellert, S.H.** (1993) *In the Wake of Chaos*, Chicago, IL: University of Chicago Press. (Discusses the philosophical implications of chaos theory, especially regarding predictability, determinism, scientific understanding and social context.)
- Smith, P.** (1991) 'The Butterfly Effect', *Proceedings of the Aristotelian Society* 91: 247-67. (Presents an argument on the implications of chaos theory for scientific prediction.)

Church's theorem and the decision problem

Church's theorem, published in 1936, states that the set of valid formulas of first-order logic is not effectively decidable: there is no method or algorithm for deciding which formulas of first-order logic are valid. Church's paper exhibited an undecidable combinatorial problem $\mathcal{P}$ and showed that $\mathcal{P}$ was representable in first-order logic. If first-order logic were decidable, $\mathcal{P}$ would also be decidable. Since $\mathcal{P}$ is undecidable, first-order logic must also be undecidable.

Church's theorem is a negative solution to the decision problem (Entscheidungsproblem), the problem of finding a method for deciding whether a given formula of first-order logic is valid, or satisfiable, or neither. The great contribution of Church (and, independently, Turing) was not merely to prove that there is no method but also to propose a mathematical definition of the notion of 'effectively solvable problem', that is, a problem solvable by means of a method or algorithm.

The period 1929-37 was very eventful for logic. It began with Gödel's proof that first-order logic is semantically complete, that is, that the set of provable formulas coincides with the set of logically valid formulas. It included the incompleteness theorems of Gödel (see [Gödel's theorems §§3-4](#)), Church's theorem showing the undecidability of first-order logic, and the statement by Church and Alan Turing of Church's thesis equating the intuitive notion of effectiveness and the mathematical notion of recursiveness. The paper by Turing, which exhibited a universal Turing machine and introduced the notion of a computer program, was the culmination of the period.

Church's theorem, published in 1936 (see Davis [1965](#)), states that the set of valid formulas of first-order logic is not effectively decidable: there is no method or algorithm for deciding which formulas of first-order logic are valid. Turing's slightly later paper proved the same result in a slightly different way. Both papers followed the same pattern. They exhibited an undecidable combinatorial problem $\mathcal{P}$ and showed that $\mathcal{P}$ was representable in first-order logic. If first-order logic were decidable, $\mathcal{P}$ would also be decidable. Since $\mathcal{P}$ is undecidable, first-order logic must also be undecidable.

Hilbert and Ackermann had formulated the *Entscheidungsproblem* (the decision problem) as the problem of finding a method for deciding whether a given formula of first-order logic is valid, or satisfiable, or neither; that is, whether all, some, or none of its possible interpretations are true. Church's theorem is a negative solution to the *Entscheidungsproblem*.

Hilbert and Ackermann had not proposed a mathematical definition of what constitutes a method, and the great contribution of Church and Turing was not merely to prove that there is no method but also to propose a mathematical definition of the notion of 'effectively solvable problem', that is, a problem solvable by means of a method or algorithm.

The technical result proved by Church only showed that the set of *provable* formulas of first-order logic is not λ -definable, that is, not definable in the λ -calculus, a logical calculus invented by Church (see [Lambda calculus](#)). To conclude that the set of *valid* formulas in first-order logic is not 'effectively' decidable, we need two additional facts. We need to equate validity in first-order logic with provability in a particular formal system, a result proved already by Gödel, but about which Church had some qualms due to the non-constructive nature of the proof. We also need to equate the mathematical notion of λ -definability (or the provably equivalent notions of recursiveness and Turing computability) with the *intuitive* notion of effective computability. This second equality is Church's thesis (see [Church's thesis](#)). Assuming Church's thesis, Church's theorem constitutes a negative solution to the problem of giving a uniform procedure for deciding which formulas of first-order logic are valid.

The particular version of first-order logic which Church used to prove his theorem included, as part of the language, the equality symbol, a constant and several function symbols. However, such a rich vocabulary is not actually necessary. Even first-order logic with a single dyadic predicate symbol is undecidable. (This is in sharp contrast to the decidability of the special case where the language consists solely of monadic predicates, a result first proved by Löwenheim ([1915](#)).)

Also, despite Church's qualms about the non-constructiveness of Gödel's completeness proof, Church's theorem is

rather sturdy. It is known that, for Gödel's completeness result to hold, one must allow formulas with non-recursive models to be considered to be satisfiable. However, even if one restricts the possible models - for example, if one confines oneself to finite models or to constructive (recursive) models - the set of valid formulas remains non-recursive, and Church's theorem remains true, although first-order logic is no longer complete relative to these restricted classes of models.

Hilbert's *Entscheidungsproblem* as formulated by Hilbert and Ackermann inquired about the decision problem for logic as such. This would also include second-order logic, whose undecidability follows of course from that of the (weaker) first-order logic. But here a slightly more complex situation prevails. First-order logic, while not decidable, is none the less semi-decidable in that we do have a procedure to enumerate the set of all valid formulas. What prevents first-order logic from being decidable is that we do not have a procedure to enumerate the complementary set of all formulas which are *not* valid, that is, of all formulas whose negations are satisfiable (have at least one true interpretation). However, second-order logic is not even semi-decidable, a fact which is already implicit in Gödel's incompleteness result for arithmetic, and which pre-dated Church's theorem by about five years (see [Gödel's theorems §1](#)). Thus, at the time that Church proved his theorem, only the question for first-order logic was really open.

Note that a positive solution to the *Entscheidungsproblem* would not have required a mathematical definition of effectiveness. It would have sufficed to present a method for deciding which formulas of first-order logic were valid and to rely on our mathematical intuition to see that the method qualified as a genuine algorithm. However, a negative solution requires us to delimit the notion of effective computability to show that validity in first-order logic lies outside it, and such a delimitation necessarily requires that the notion of effectiveness be given a mathematical definition, or at least be mathematically circumscribed.

Church himself did not provide in his paper an independent definition of the intuitive notion of effectiveness. Rather he proposed that his technical notion, λ -definability, be taken to be a definition of the intuitive notion of effectiveness. A more substantial analysis was provided by Alan Turing in his paper (see [Davis 1965](#)). Turing analysed the possible mental algorithms that a human being could carry out and claimed that his machines, Turing machines, could also carry them out. A crucial element in Turing's argument was the claim that a human computer (sic) would only be able to carry out algorithms which utilized a finite number of distinct symbols, and that while he was carrying out an algorithm, his mind could only occupy a finite number of distinct states. Turing's argument for the first claim is that if the process of calculation used infinitely many basic symbols, then some of these symbols would be very close to others and confusion could not be avoided. He remarks that the same also holds of states of mind, but offers no argument. (See [Turing machines](#).) Gödel, in fact, describes Turing's remarks in this context as a 'philosophical error'. If we do accept Turing's justification that effectiveness equals Turing computability, then since λ -definability and Turing computability are provably equivalent, Turing's analysis also provides substantial ammunition for Church's own claim.

See also: [Church's thesis](#); [Logical and mathematical terms, glossary of](#)

ROHIT PARIKH

References and further reading

- Davis, M.** (ed.) (1965) *The Undecidable: Basic Papers on Undecidable Propositions, Unsolvability Problems and Computable Functions*, Hewlett, NY: Raven Press. (Includes a corrected version of Church's original paper, and Turing's article.)
- Gödel, K.** (1990) 'Some Remarks on the Undecidability Results', in *Collected Works*, ed. S. Feferman *et al.*, New York and Oxford: Oxford University Press, vol. 2, 305-6. (An excellent account of Gödel's own views.)
- Herken, R.** (ed.) (1988) *The Universal Turing Machine: A Half-Century Survey*, Oxford: Oxford University Press. (An excellent collection of papers by contemporary researchers; see especially R. Gandy, 'The Confluence of Ideas in 1936', 55-111.)
- Löwenheim, L.** (1915) 'Über Möglichkeiten im Relativkalkül', *Mathematische Annalen* 76: 447-70; trans. S. Bauer-Mengelberg, 'On Possibilities in the Calculus of Relatives', in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967, 232-51. (Decidability of the special case of first-order logic with a language consisting solely of monadic predicates.)

Church's thesis

An algorithm or mechanical procedure A is said to 'compute' a function f if, for any n in the domain of f , when given n as input, A eventually produces fn as output. A function is 'computable' if there is an algorithm that computes it. A set S is 'decidable' if there is an algorithm that decides membership in S : if, given any appropriate n as input, the algorithm would output 'yes' if $n \in S$, and 'no' if $n \notin S$. The notions of 'algorithm', 'computable' and 'decidable' are informal (or pre-formal) in that they have meaning independently of, and prior to, attempts at rigorous formulation.

Church's thesis, first proposed by Alonzo Church in a paper published in 1936, is the assertion that a function is computable if and only if it is recursive: 'We now define the notion...of an effectively calculable function...by identifying it with the notion of a recursive function....' Independently, Alan Turing argued that a function is computable if and only if there is a Turing machine that computes it; and he showed that a function is Turing-computable if and only if it is recursive.

Church's thesis is widely accepted today. Since an algorithm can be 'read off' a recursive derivation, every recursive function is computable. Three types of 'evidence' have been cited for the converse. First, every algorithm that has been examined has been shown to compute a recursive function. Second, Turing, Church and others provided analyses of the moves available to a person following a mechanical procedure, arguing that everything can be simulated by a Turing machine, a recursive derivation, and so on. The third consideration is 'confluence'. Several different characterizations, developed more or less independently, have been shown to be coextensive, suggesting that all of them are on target. The list includes recursiveness, Turing computability, Herbrand-Gödel derivability, λ -definability and Markov algorithm computability.

1 Computability and decidability

The algorithms or mechanical procedures envisaged here must determine what is to be done at each step, involving no creativity or intuition on the part of the agent (other than an ability to recognize symbols). The paradigm is a recipe for routine calculation, or a computer program.

An algorithm (or mechanical procedure) A 'computes' a function f if, for any n in the domain of f , when given n as input, A produces fn as output. A function is 'computable' if there is an algorithm that computes it. (See [Computability theory §2](#).) There is no fixed limit on the amount of materials or memory required for a computation, nor on the number of steps an algorithm might take. If a function were computed only by an algorithm whose execution required more memory than there are particles in the solar system, it would not be capable of computation in any realistic sense, yet it would still be computable. What is required is that the algorithm have only finitely many steps. A set S is 'decidable' if there is an algorithm that 'decides' membership in S : if, given any appropriate n as input, the algorithm outputs 'yes' if $n \in S$, and 'no' if $n \notin S$.

Strictly speaking, computability should apply only to functions on things like marks on paper and magnetic media, since these are the items that can be manipulated by computers and algorithms. It is common to take computability to apply to functions on 'strings', which are the abstract forms of sequences of characters. For example, the string 'abc23' is the form of all actual and possible concrete inscriptions of these three letters and two numerals. It is also common to think of computability as applying to functions on natural numbers, via canonical notation.

The root meaning of 'computability' would be something like 'capable of computation', and the root meaning of 'decidability', something like 'capable of decision'. Thus, these are modal notions. Computability is also an 'extensional' notion: it applies to functions themselves, not to descriptions or presentations of functions. Moreover, the quantifier 'there is an algorithm' does not involve (*de re*) knowledge of an algorithm (see [De re/de dicto §1](#)). For example, let $fn = 1$ if there is a string of at least n consecutive 5's in the decimal expansion of π , and $fn = 0$ otherwise. No one knows of any algorithm A such that A computes f . Nevertheless, the function f is computable. If there are indefinitely long strings of consecutive 5's in the expansion of π , then f is computed by an algorithm that produces '1' no matter the input. Otherwise, let N be the longest string of consecutive 5's in the expansion. Then f is computed by an algorithm that outputs '1' for all inputs less than or equal to N , and outputs '0' for all other inputs. In all cases, then, f is computable.

The notions of recursiveness (see [Computability theory §1](#)) and λ -definability (see [Lambda calculus §4](#)) arose as part of various foundational programmes earlier in the twentieth century. Church and his students became interested in the extension of the λ -definable functions. They set themselves the task of investigating various (computable) functions, to see if they could be shown to be λ -definable. Shortly after Kleene proved that the predecessor function is λ -definable, Church proposed his thesis: that a function is computable if and only if it is recursive. He published it after Kleene established the connection with recursiveness.

Although algorithms have been constructed since antiquity, it is curious that no one attempted a precise characterization of computability *per se* before the mid-1930s, at which time a number of more or less independent characterizations appeared almost simultaneously. One motivation was the desire, or need, to establish *negative* results about computability and decidability. Notice that the straightforward way to show that a given function is computable is to give an algorithm. This does not require a rigorous definition of computability, since most theorists can recognize an algorithm when they see one. To establish that a given function is *not* computable, on the other hand, one needs to show that *no* algorithm computes the function. This requires the formulation of a property shared by all algorithms or all computable functions. The straightforward route to such a property is a careful characterization of computability. At the time, the construction of algorithms for various purposes was a central item on the agenda of mathematical logic (see [Hilbert's programme and formalism](#)). Virtually all of the original characterizations of computability include theorems that certain functions are not computable or that certain sets are not decidable.

A related motivation was the desire to understand the extent of Gödel's incompleteness theorem. The original proof only applied to a particular system, but it is clear that the proof could be generalized to any common axiomatization of arithmetic. This, and the work on Hilbert's programme, led to the problem of characterizing just what a formal deductive system is. One requirement is that the syntax should be 'effective', in the sense that the collections of well-formed formulas, axioms and rules of inference should all be decidable. The natural generalization of the incompleteness theorem is that there is no sufficiently rich, effective and complete deductive system for arithmetic. As above, to establish this generalization, one needs a grasp of the limits of algorithms, and this came with the characterization of computability (see [Gödel's theorems](#); [Formal languages and systems](#)).

The modern notion of a function as an arbitrary correspondence between sets has emerged only towards the end of the twentieth century. In the past, the very idea of a 'non-computable' function might have been questioned, as it is today by constructivists and intuitionists. It might be added that from the constructivist perspective, Church's thesis has a sharp formulation. For a constructivist, a formula of the form $\forall x \exists! y \Phi(x, y)$ states that, given any x , *one can find* a unique y such that $\Phi(x, y)$ holds. This amounts to an assertion that a certain function is computable. Thus, Church's thesis is the scheme that if $\forall x \exists! y \Phi(x, y)$ then there is a *recursive* function f such that $\forall x \Phi(x, f(x))$. In other words, for a constructivist, Church's thesis is a claim that *all* number-theoretic functions are recursive. Among intuitionists and constructivists, Church's thesis is controversial (see [Intuitionism](#); [Constructivism in mathematics](#); Beeson 1985; McCarty 1987).

As noted, however, there is a consensus among (non-constructivist) mathematicians and logicians that Church's thesis is correct. Within recursive function theory, a technique has been developed, sometimes called 'argument by Church's thesis', wherein one concludes that a function is recursive, or that a set is decidable, just because there is an algorithm for it. For example, for each natural number n , let $g(n)$ be the sum of the divisors of n . Since it is routine to give an algorithm to factor a number and add its divisors, one can conclude, by Church's thesis, that g is recursive. One does not have to deal with the details of recursive derivations, Turing machines and so on. By avoiding tedious detail, the perspective is broadened, leading to greater insight (see Rogers 1967).

2 What is Church's thesis?

Despite the consensus on the correctness of Church's thesis, there is not much agreement over what the correctness amounts to. What does Church's thesis mean? Does it have a determinate truth-value? Can it be known and, if so, how?

Computability is a pragmatic, pre-theoretic property of functions. It relates somehow to human abilities, suitably idealized. As such, Church's thesis is an important case study for the philosophical problem of accounting for the relationship between mathematics and material or human reality. Some authors have gone so far as to call

computability a 'psychological' concept, since it relates to competence in grasping and executing procedures. One of the founders of computability, Emil Post, wrote that

for full generality a complete analysis would have to be made of all the possible ways in which the human mind could set up finite processes. . . we have to do with a certain activity of the human mind as situated in the universe. As activity, this logico-mathematical process has certain temporal properties; as situated in the universe it has certain spatial properties.

(Davis 1965: 408-19)

This passage also notes the aforementioned modal nature of computability. Notice, however, that recursiveness, Turing computability, λ -definability and so on are at least *prima facie* non-modal. They are defined in terms of quantifiers ranging over abstract entities, such as natural numbers, strings on a finite alphabet and so on. Thus, Church's thesis identifies the extension of an idealized, pragmatic, modal property of functions with the extension of a formal, rigorously defined, *prima facie* non-modal arithmetic property. On the surface, then, Church's thesis proposes an exchange of modality for ontology. How are such proposals to be evaluated?

A closely related matter concerns vagueness. Typically, pragmatic, modal notions do not have sharp boundaries. There are, for example, borderline cases of 'climbable mountains', 'liftable weights' and 'teachable students'. Clearly, there are also borderline cases of functions 'capable of computation', if we do not idealize. Slogans such as 'ignoring limits on memory and materials' sharpen the notion, but it is still not clear that idealized computability, as defined above, is an unequivocal property with *completely* precise boundaries. Recursiveness and Turing computability, on the other hand, do appear to have precise boundaries. *Prima facie*, then, Church's thesis identifies a vague notion with a precise one.

Clearly, a vague property cannot *exactly* coincide with a precise one. So, if computability is in fact vague, Church's thesis does not literally have a determinate truth-value or else it is false. Church's thesis might then be thought of as a proposal that recursiveness be *substituted* for computability, for certain purposes. It would be similar to a pronouncement such as 'For the purposes of this form, a resident is defined to be...'.

The value of 'definitions' or 'precisifications' of vague terms depends on the purposes at hand. Given the extensive idealizations, Church's thesis seems best suited for establishing *negative* results about computability. If one shows, for example, that a function is not recursive, that is conclusive reason to hold that it is not 'capable of computation': one should give up trying to compute it. Suppose, on the other hand, that a function has been shown to be recursive. By itself, this gives no information on the feasibility of an algorithm, and thus it does not establish that the function can be calculated, in any realistic sense. Moreover, it is of little value to anyone interested in actually computing the function. For this purpose, there are other, less idealized models, such as 'finite state' computability, 'push down' computability, primitive recursiveness and computability in polynomial time or space (see **Complexity, computational**).

A different interpretation of Church's thesis is that even though computability is a pre-theoretic notion, it is nevertheless a precise, mathematical property of functions. One might hold, for example, that there are determinate mathematical structures underlying idealized human abilities to calculate, write programs, or even think. Understood this way, Church's thesis states that this property is coextensive with another precise mathematical notion, recursiveness. Thus, Church's thesis has a truth-value. The 'arguments from confluence' support this interpretation. The fact that a number of *prima facie* different attempts to characterize computability lead to the same extension suggests that recursiveness, Turing computability and so on delimit the boundaries of a 'natural kind' underlying computation.

3 Epistemic status

The Goldbach conjecture (that every even number greater than 4 is the sum of two primes) and the 'twin prime' conjecture (that there are infinitely many numbers n such that n and $n + 2$ are both prime) are long-standing open problems in number theory. They can be settled, if at all, by mathematical proof. It is widely held that Church's thesis is not like this. Its evidence is not mathematical. Because of the connection with human abilities, and perhaps because of the supposed vagueness, Church's thesis has been called an 'empirical' or 'quasi-empirical' matter.

As noted above, one type of evidence cited for Church's thesis is that every algorithm that has been investigated has been shown to compute a recursive function. This is consonant with the quasi-empirical nature of Church's thesis (on a certain view of scientific hypotheses). The intuitive ability to recognize algorithms is analogous to 'observation', and showing that an algorithm computes a recursive function amounts to 'confirmation' of Church's thesis (see [Confirmation theory](#)). This type of evidence was prominent in the thought of Church and Kleene.

Some theorists, on the other hand, do regard Church's thesis as a mathematical matter, capable of whatever epistemic standards are applicable there, and some regard the thesis as proved. Robin Gandy (1988), for example, suggests that Turing's direct argument that every algorithm can be simulated on a Turing machine does not merely provide evidence for an inductive, scientific hypothesis; rather, it 'proves a theorem'. Turing provides a careful analysis of a human executing an algorithm, showing how each 'step' can be simulated on a Turing machine. There are lemmas that the number of symbols and the number of 'states of mind' must be finite. Gandy regards the analysis to be as convincing as typical mathematical work. Apparently, Gödel was also convinced by Turing's arguments, and not by the 'quasi-empirical' evidence (see Davis 1982).

On these views, the pragmatic nature of computability, and its potential vagueness, do not automatically disqualify Church's thesis from mathematics. This is consistent with a widely held position that there is no sharp boundary between mathematics and empirical science (see [Duhem, P.M.M.](#); [Quine, W.V.O.](#)). Mathematics has its roots in everyday life, and cannot be severed from it. Even a cursory look at the history of the notions underlying recursiveness reveals ambivalence, ambiguity and vagueness, especially in the notion of 'set' (see Mendelson 1990). Yet no one doubts that sets and recursive functions belong to mathematics. According to the view at hand, computability is no different.

4 Other theses

There are similar 'theses' in the history of mathematics and science. Consideration of these may shed some light on the philosophical issues here.

The notion of a grammatical sentence of a natural language is clearly pragmatic, modal and at least *prima facie* vague. This is contrasted with formal, mathematical characterizations of grammaticality, as proposed by linguists. Unlike Church's thesis, however, there is no consensus on one of these grammars.

Closer to home, the notion of 'validity', or 'logical consequence', is *prima facie* pragmatic and modal, if not vague. In typical logic textbooks, validity is related to inference abilities, and an argument is said to be valid if it is not possible for the premises to be true and the conclusion false. Yet, when one reads on, validity is formally defined in terms of models, which are mathematical constructions on sets. Thus, we have another exchange of modality and ontology. The identification of the pre-formal validity with model-theoretic consequence may be called 'Tarski's thesis'. Notice that there is a consideration of confluence here. Gödel's completeness theorem shows that for first-order languages, the model-theoretic notion of consequence coincides in extension with a deductive one (see [Consequence, conceptions of](#); [Model theory](#); [Proof theory](#)).

Within more traditional mathematics, consider notions such as 'area', 'volume' and 'continuity'. While these do not appear pragmatic and modal, they did have clear meanings before their rigorous formulations in analysis and measure theory. Moreover, the notions do have uses in everyday and scientific life and there seem to be borderline cases of the pre-theoretic notions. In modern mathematics, however, these notions all have rigorous, precise definitions. The various identifications with the pre-theoretic notions may be called Euclid's thesis, Lebesgue's thesis, Weierstrass' thesis, Cauchy's thesis, and so on. The reader may decide the extent to which these theses have determinate truth-values, whether they are mathematical or 'quasi-empirical', and whether they have been (or can be) established with characteristic mathematical rigour.

See also: [Computability theory](#); [Complexity, computational](#); [Computability and information](#); [Logic machines and diagrams](#); [Logical and mathematical terms, glossary of](#)

STEWART SHAPIRO

References and further reading

Beeson, M.J. (1985) *Foundations of Constructive Mathematics*, New York: Springer.(Includes an account of

Church's thesis from a constructivist perspective.)

- Davis, M.** (ed.) (1965) *The Undecidable: Basic Papers on Undecidable Propositions, Unsolvability Problems and Computable Functions*, Hewlett, NY: Raven Press. (Original papers, or English translations, by Gödel, Church, Turing, Rosser, Kleene and Post; includes some items not published elsewhere.)
- Davis, M.** (1982) 'Why Gödel Didn't Have Church's Thesis', *Information and Control* 54: 3-24. (Careful historical discussion of what convinced some of the major players of Church's thesis.)
- Gandy, R.** (1988) 'The Confluence of Ideas in 1936', in R. Herken (ed.) *The Universal Turing Machine: A Half-Century Survey*, New York: Oxford University Press, 55-111. (Extensive historical study of the development of computability.)
- Kleene, S.** (1979) 'Origins of Recursive Function Theory', in *Twentieth Annual Symposium on Foundations of Computer Science*, New York: IEEE, 371-82. (Reminiscent account, by one of the founders of recursive function theory; good source on the development of the λ -calculus and on Church's thought on logic and computability.)
- Kleene, S.** (1987) 'Reflections on Church's Thesis', *Notre Dame Journal of Formal Logic* 28: 490-8. (Philosophical discussion of Church's thesis.)
- McCarty, C.** (1987) 'Variations on a Thesis: Intuitionism and Computability', *Notre Dame Journal of Formal Logic* 28: 536-80. (An intuitionistic view of Church's thesis.)
- Mendelson, E.** (1990) 'Second Thoughts about Church's Thesis and Mathematical Proofs', *Journal of Philosophy* 87: 225-33. (Argument that Church's thesis is provable.)
- Myhill, J.** (1952) 'Some Philosophical Implications of Mathematical Logic', *Review of Metaphysics* 6: 165-98. (Philosophical account of the argument that computability is partly psychological.)
- Rogers, H.** (1967) *Theory of Recursive Functions and Effective Computability*, New York: McGraw-Hill. (Comprehensive mathematical development, based on 'argument by Church's thesis'.)
- Shapiro, S.** (1981) 'Understanding Church's Thesis', *Journal of Philosophical Logic* 10: 353-65. (Discussion of the philosophical status of Church's thesis.)
- Shapiro, S.** (1983) 'Remarks on the Development of Computability', *History and Philosophy of Logic* 4: 203-20. (Historical account.)
- Shapiro, S.** (1993) 'Understanding Church's Thesis, Again', *Acta Analytica* 11: 59-77. (A further discussion of the philosophical status of Church's thesis, including its modal status, its vagueness and the view that Church's thesis is a mathematical matter.)
- Sieg, W.** (1994) 'Mechanical Procedures and Mathematical Experience', in A. George (ed.) *Mathematics and Mind*, Oxford: Oxford University Press, 71-117. (Philosophical account of Church's thesis and its relationships with mechanism.)

Combinatory logic

Combinatory logic comprises a battery of formalisms for expressing and studying properties of operations constitutive to contemporary logic and its applications. The sole syntactic category in combinatory logic is that of the applicative term. Closed terms are called ‘combinators’; there is no binding of variables. Systems containing the basic combinators S and K exhibit the crucial property of combinatorial completeness: every routine expressible in the system can be captured by a term composed of these two combinators alone. Combinatory logic is a close relative of Church’s lambda calculus. M. Schönfinkel first introduced and defined basic combinators in 1920 in assaying foundations for mathematics that avoid bound variables and take operations, rather than sets, as fundamental. H. Curry later rediscovered the combinators (and coined the term ‘combinatory logic’) independently of Schönfinkel. Curry constructed various formal systems for combinatory logic and, throughout most of the subject’s history, was the central figure in the research. In 1969, D. Scott succeeded in constructing set-theoretic, functional models for the lambda calculus and combinatory logic. Since then semantic studies of combinatory systems, together with research on their applications to computer science and further development as foundational systems, have dominated the field.

1 Combinators and the elimination of variables

In his *On the Building Blocks of Mathematical Logic* (1924), M. Schönfinkel sought to surpass Whitehead and Russell’s *Principia Mathematica* and Frege’s *Grundgesetze* in isolating notions absolutely fundamental to mathematical logic. In addition to employing a Sheffer stroke-like operator for predicate calculus (Sheffer (1913) showed that all truth-functional connectives could be expressed in terms of ‘not...and’), Schönfinkel proposed abandoning the type-stratified logical universe of *Principia Mathematica* in favour of one consisting entirely of type-free functions, mathematical operations presupposed by all others, unlimited in their allowed inputs and, hence, self-applicable. Moreover, he looked to reduce the number of all logical symbols (of both first- and higher-order logic) to three: the stroke-like function, a constant function and a ‘fusion’ function. These were the building blocks from which Schönfinkel intended to analyse logical expressions as compounds of basic operations. In this way, the ultimate operations on which mathematical logic depends - operations usually left implicit and unformalized - were to be revealed to the foundational gaze. As a bonus, this universal analysis was to require a wholesale elimination from logic of the variable.

For a simple example of variable elimination, consider a binary operation $\circ$ which is ‘idempotent’; that is, for all x , $x \circ x = x$. To render this relation without bound variable x , one introduces operation symbols or, to use terminology due to H. Curry, ‘combinators’. In this case, let O and I be such that $Ox = x \circ x$ and $Ix = x$. Then, to claim idempotency for $\circ$, one writes merely $O = I$. The variable x becomes redundant, for, if we apply both sides of the combinator equation to x , we return the original: $x \circ x = Ox = Ix = x$. Combinators then serve to express general laws, such as those of function application, without bound variables, and knotty issues of substitutability and illicit variable capture are avoided.

In general, a combinator is a variable-free name for an operation (on operations) defined using an equation. A number of combinators are standard and bear conventional names. Some of the most common are S , K , I , B and W , defined as follows: $Sxyz = xz(yz)$, $Kxy = x$, $Ix = x$, $Bxyz = x(yz)$ and $Wxy = xyy$. Schönfinkel, in effect, introduced combinators S and K for the operations he called ‘fusion’ (*Verschmelzung*) and ‘constant’ (*Konstanz*). S performs a generalized composition operation and K , when applied to x , yields Kx , the constantly x -valued function of y . I is a universal identity function. B denotes conventional composition of functions, since Bfg , applied to x , gives $f(gx)$. Finally, W realizes diagonalization: Wf , applied to x , yields $fx x$. If we think of f as a binary function, Wf is the unary function which is f with arguments identified. (Incidentally, here we are exploiting a device known to Schönfinkel, as well as to Frege, but now called ‘currying’, after Curry. By that device, combinators denote unary operations but achieve the effect of n -ary operations ($n > 1$) when iterated.)

2 Combinatory logic

Curry (1900-82), a student of Hilbert who rediscovered the combinators independently of Schönfinkel, devised the first formal systems for (what he termed) ‘combinatory logic’. In Curry’s systems, the fundamental syntactic

category is that of (combinatory) term, where basic terms are either variables or constants. Compound terms denote applications: if t and u are terms, then so is tu , expressing the application of t to u . Among the constants are the basic combinators K and S . Any term built up entirely from K and S is also called a combinator.

As with the lambda calculus, we can define appropriate notions of reduction, normal form and equality for combinatory terms (see [Lambda calculus](#)). The ‘reduction’ of one combinatory term to another is conceived by analogy with simplifying computation, for example, the reduction of $(3 + 6) \times (10 - 2)$ to 72 by calculation. A ‘normal form’ is a term which can be simplified no further. Terms t and u are equal whenever they can be linked by a finite sequence of terms $t_0, t_1, \dots, t_n$ such that, for every k and l , either t_k reduces to t_l or conversely.

Since operations may be viewed both extensionally and non-extensionally, there are extensional (‘strong’) and non-extensional (‘weak’) conceptions of reduction and normality, on either of which Kxy reduces to x and $Sxyz$ to $xz(yz)$. Both kinds of reduction are univocal in that, if t reduces to a normal form, it does so uniquely. This is guaranteed by the ‘Church-Rosser theorem’, which states that, if t reduces to u and to v , then there is a term w to which u and v both reduce. As in the lambda calculus, select combinators can serve as numerals and an arithmetic can be so developed that every partial recursive function is representable within combinatory logic.

It is readily seen that some combinators are definable from others. For example, I is recovered as SKK , since, for any x , $SKKx = Kx(Kx) = x = Ix$. The principle here is general: S and K comprise a ‘complete’ set of combinators in that every other combinator is derivable from these alone. Indeed, let t be any term, simple or complex, containing only variables from a finite set, say, x, y, z . Then there is a combinator, denoted $[x, y, z]t$, called ‘the abstraction of t with respect to x, y and z ’, containing no free variables and constructed entirely from S and K such that $(([x, y, z]tx)y)z$ reduces (weakly) to t . In effect, Schönfinkel was the first to recognize the completeness of combinatory logic.

The notation ‘ $[x, y, z]t$ ’ is entirely metatheoretic; it denotes a term containing neither free nor bound variables and connotes that the abstraction $[x, y, z]$ -operation plays a role in combinatory logic analogous to that played in the lambda calculus by the $\lambda x \lambda y \lambda z$ abstractor (see [Lambda calculus](#)). This loose analogy is made precise by defining functions which inject the set of combinatory terms into those of the lambda calculus so that relevant equality relations coincide. There are metatheoretic functions Λ and C , where Λ maps combinatory terms into lambda terms and C the reverse, so that, for combinatory terms t and u , $C\Lambda(t)$ is the same term as t and, as far as extensional equality is concerned, $t = u$ in combinatory logic if and only if, in the lambda calculus, $\Lambda(t)$ equals $\Lambda(u)$. (The non-extensional versions of the respective calculi are also comparable but the matter is more complex.) The function C employs abstraction in dealing with lambda terms: $C(\lambda x \lambda y. t)$, for example, is $[x, y]C(t)$.

Intuitively, a model for (non-extensional) combinatory logic would be a universe D with a binary operation $\bullet$ - representing application - and containing elements k and s which behave, with respect to $\bullet$, as K and S do. It is easy to verify that such a model satisfies all the equalities of the weak or non-extensional combinatory logic. However, combinatory logic was originally conceived as a doctrine of operations or functions and real difficulties arise in rendering these functions in set-theoretic terms. For one thing, combinatory logic allows self-application: the combinator SS is perfectly well-formed. But no standard set-theoretic function can be a member of its own domain. Moreover, every combinator, unlike every set-theoretic function, admits fixed points. Curry discovered a ‘paradoxical’ combinator Y , which may be defined $WS(BWB)$, that computes a fixed point for any term: if t is a term, Yt is equal to $t(Yt)$.

In 1969, D. Scott (1976) was the first to overcome these obstacles and give set-theoretic constructions of models (in the above sense) for combinatory logic in which the terms name functions. A key idea was to let combinatory terms name functions which are not arbitrary but continuous. Here, a continuous function on a partial order with limits preserves the order and the limits which exist. Scott’s original construction was quite complex; models less complicated to build were soon forthcoming. These semantic breakthroughs inaugurated a period of intense research on combinatory logic and the lambda calculus, stimulating numerous applications of both to computer science.

3 Combinatory logic as a foundation for mathematics

Schönfinkel’s original goal, which Curry nurtured and pursued under the rubric ‘illative combinatory logic’, was

the construction of a comprehensive system for the foundation of mathematics free of types and bound variables, and highlighting those operations essential to logical thought. To be comprehensive, no significant notion was to be banished to the metatheory. Other than the assertion sign, every concept which bears on the system was to be amenable to combination with combinators inside the system.

The strongest of Curry's early illative systems, which appeared in print in 1934, proved excessively comprehensive: Church's students S. Kleene and J. Rosser demonstrated its inconsistency in 1935. Curry himself refined the core of their argument, obtaining the elegant 'Curry's paradox', which shows plainly that even a simple logic of implication cannot be developed in the presence of combinatorial completeness. Curry and his co-workers then sought to avoid paradox by inventing systems to circumscribe, within the class of legal terms, canonical terms standing for propositions. Unfortunately, consistent versions of these systems were too weak to support a comprehensive mathematics. Throughout the middle and late twentieth century, a number of logicians contributed significantly to achieving the goal of Curry and Schönfinkel (a task known as 'Curry's programme' - see [Seldin 1980](#)) for illative systems, among them P. Aczel, M. Bunder, S. Feferman, F. Fitch, J. Hindley, D. Scott and J. Seldin. In 1983, M. Bunder published his system HOPC ('Higher-Order Predicate Calculus'), which is an illative system with internal typing and in which a significant portion of Zermelo-Fraenkel set theory can be interpreted.

See also: [Logical and mathematical terms, glossary of](#)

DAVID CHARLES McCARTY

References and further reading

- Barendregt, H.** (1981) *The Lambda Calculus: Its Syntax and Semantics*, Amsterdam: North Holland, 2nd edn, 1984. (A bible of lambda calculus research. Authoritative and detailed.)
- Curry, H.** (1934) 'Functionality in Combinatory Logic', *Proceedings of the National Academy of Sciences of the USA* 20: 584-90. (An article crucial to later research in illative combinatory logic.)
- Curry, H. and Feys, R.** (1958) *Combinatory Logic*, vol. 1, Amsterdam: North Holland. (Once the premier work on the subject; it remains an excellent reference.)
- Hindley, J. and Seldin, J.** (1986) *Introduction to Combinators and λ -Calculus*, London Mathematical Society Student Texts no. 1, Cambridge: Cambridge University Press. (An excellent introduction, highly recommended for beginners.)
- Kleene, S. and Rosser, J.** (1935) 'The Inconsistency of Certain Formal Logics', *Annals of Mathematics* 36: 630-6. (A technical application of gödelization to proving the inconsistency of early systems of Church and Curry.)
- Schönfinkel, M.** (1924) 'Über die Bausteine der mathematischen Logik', *Mathematische Annalen* 92: 305-16; trans. *On the Building Blocks of Mathematical Logic*, ed. H. Behmann; repr. in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967, 355-66. (From a 1920 presentation to the Mathematical Society at Göttingen. Introduction in the 1967 version by W.V. Quine.)
- Scott, D.** (1976) 'Data Types as Lattices', *SIAM Journal of Computing* 5: 522-87. (A characteristically charming exposition of results by Scott, Plotkin and others on the graph model.)
- Seldin, J.** (1980) 'Curry's Program', in J. Hindley and J. Seldin (eds) *To H.B. Curry: Essays on Combinatory Logic, Lambda Calculus and Formalism*, New York: Academic Press, 3-34. (An authoritative presentation of Curry's vision for combinatory logic.)
- Sheffer, H.M.** (1913) 'A Set of Five Independent Postulates for Boolean Algebras, With Application To Logical Constants', *Transactions of the American Mathematical Society* 14: 481-8. (The first publication of the fact that there is a single connective that is expressively complete for sentential logic.)
- Stenlund, S.** (1972) *Combinators, λ -Terms and Proof Theory*, Dordrecht: Reidel. (A laudably concise introduction to combinatory logic and its role in proof theory.)

Common-sense reasoning, theories of

The task of formalizing common-sense reasoning within a logical framework can be viewed as an extension of the programme of formalizing mathematical and scientific reasoning that has occupied philosophers throughout much of the twentieth century. The most significant progress in applying logical techniques to the study of common-sense reasoning has been made, however, not by philosophers, but by researchers in artificial intelligence, and the logical study of common-sense reasoning is now a recognized sub-field of that discipline.

The work involved in this area is similar to what one finds in philosophical logic, but it tends to be more detailed, since the ultimate goal is to encode the information that would actually be needed to drive a reasoning agent. Still, the formal study of common-sense reasoning is not just a matter of applied logic, but has led to theoretical advances within logic itself. The most important of these is the development of a new field of 'non-monotonic' logic, in which the conclusions supported by a set of premises might have to be withdrawn as the premise set is supplemented with new information.

The formal study of common-sense reasoning is a field in which the concerns of philosophy and artificial intelligence converge. From a philosophical point of view, it is motivated by a popular approach to the philosophy of mind - known as the 'language of thought' hypothesis - that postulates a domain of mental representations as the primary bearers of meaning, and then analyses cognition as the rule-governed manipulation of these representations. In order for an approach along these lines to be useful, some account must be provided of the structure of the internal representations, and of the rules governing their manipulation (see [Language of thought](#)).

These issues have been studied extensively within cognitive psychology, of course. Here, the idea is to focus on an existing system capable of intelligent behaviour in a wide range of circumstances - a human being - and to attempt to infer the symbolic processes underlying this intelligence. But the issues are studied also within the area of artificial intelligence, and from a perspective that offers a different kind of illumination: the goal here is to build an intelligent system, rather than attempting to discover the structure of a system that already exists. Although the performance exhibited by the artificial systems designed to date is significantly less impressive than that of humans when evaluated across a broad range of circumstances, the symbolic processes underlying the behaviour of these systems are at least well understood.

Among the various research methodologies that have emerged within artificial intelligence, one of the most powerful is the logic-based approach first advanced by John McCarthy; and it is primarily this approach that has motivated the formal study of common-sense reasoning. In fact, the logic-based approach in artificial intelligence has much in common with the project of formalizing mathematical and scientific reasoning that has been carried on throughout the twentieth century. Within artificial intelligence, the idea is that much of the knowledge necessary for achieving intelligent action even in everyday situations can be represented in a machine through the formulas of some logical language, and that the reasoning tasks underlying this intelligence can then be accomplished by means of logical deduction.

Both the idea of using logic as an underlying representation language for artificial intelligence and the emphasis on formalizing common-sense knowledge are present even in McCarthy's very early work (for example, McCarthy 1959), but the project receives its clearest articulation in a paper jointly authored by McCarthy and Patrick Hayes (1969), which isolates the task of defining 'a naïve, common-sense view of the world precisely enough to programme a computer to act accordingly'.

Much of the work now produced within artificial intelligence on the topic of formalizing common-sense reasoning is similar to what one finds in philosophical logic - attempts to adapt or generalize logical techniques to apply to some new area - and there is now a good deal of interaction between these two fields. Still, the research that takes place within artificial intelligence has a somewhat different character from standard philosophical logic. For one thing, the matter of implementation is always in the background, and sometimes in the foreground; but even apart from that, the studies generated within artificial intelligence are often focused on very detailed representational problems that would seem peculiar to a philosopher. It would not be unusual, for example, to find a research project in artificial intelligence with the goal of representing what an agent would have to know about the objects

in their kitchen in order to prepare a meal.

This kind of detailed work is necessary, of course, since the ultimate goal is to encode information in such a way that it could actually be used to drive a reasoning agent. Still, it is not as if the project of providing a logical representation of common-sense knowledge were simply a matter of applied logic. In fact, the problems presented by the task of representing this kind of information explicitly have led to a number of theoretical advances within logic itself: the most important of these is the development of the new field of ‘non-monotonic’ logic.

In most standard logics, the addition of new information to a given set of premises might lead us to draw new conclusions, but never to withdraw conclusions already reached. The set of consequences of a given set of premises is thus said to grow monotonically as the premise set grows: the consequence set can only increase, never decrease, as the premise set is supplemented with new information. A non-monotonic logic is simply a logic in which this property fails - a logic in which the addition of new information to a given premise set might force us to retract some conclusion drawn from the original set of premises (see [Non-monotonic logic](#)).

The study of non-monotonic logics within artificial intelligence was motivated by the realization that much of our common-sense knowledge concerns defeasible information - generalizations subject to exceptions, such as ‘Birds fly’ or ‘Things remain where you put them’. Of course, philosophers had always known that these defeasible generalizations could not be represented naturally in ordinary logic: the statement ‘Birds fly’, for example, cannot be represented by a formula of the form $\forall x(Bx \supset Fx)$. But it was not until the practical need arose for reasoning with defeasible statements such as these that serious attention was focused on the problem; and it then became apparent that the appropriate notion of defeasible consequence would have to be non-monotonic.

To take a standard example, given only the information that Tweety is a bird and that birds tend to fly, it is natural to conclude defeasibly that Tweety flies. But if this premise set were supplemented, consistently, with the additional information that Tweety does not fly (perhaps Tweety is a penguin), it would then seem reasonable to withdraw our initial conclusion.

The monotonicity property flows from assumptions that are deeply rooted in both the proof theory and the semantics of most ordinary logics. From a proof-theoretic point of view, this property follows from the fact that a proof based on a set of premises also counts as a proof based on an expansion of the same set of premises; from a semantic point of view, the property results from the assumption that the models of a set of premises are models also of any of its subsets. Because the features underlying the monotonicity property are so basic to the conception of most standard logical systems, researchers concerned with non-monotonic reasoning have been led to explore fundamentally new ideas in both proof theory and semantics.

The first proof-theoretic treatment of non-monotonic reasoning is found in Raymond Reiter’s default logic (1980), which supplements ordinary logic with new rules of inference, known as ‘default rules’. The semantic approach to non-monotonic reasoning was first explored in McCarthy’s own theory of circumscription (1980).

Because of its intrinsic interest and practical importance, the study of non-monotonic logic has grown into a significant area of research. The fixed-point and minimal model approaches pioneered by Reiter and McCarthy are still the best-known and most widely applied techniques in non-monotonic reasoning, but a number of other approaches have been explored as well; a series of articles surveying these different approaches can be found in Gabbay *et al.* (1994). In recent years, the techniques developed within the field of non-monotonic logic have begun to find applications also to a number of philosophical issues, such as the understanding of *ceteris paribus* clauses in scientific generalizations and the formalization of *prima facie* obligations.

See also: [Artificial intelligence](#)

JOHN HORTY

References and further reading

- Davis, E.** (1990) *Representations of Commonsense Knowledge*, San Mateo, CA: Morgan-Kaufmann. (A textbook in the field of formalizing commonsense reasoning.)
- Gabbay, D., Hogger, C.J. and Robinson, J.A.** (eds) (1994) *Handbook of Logic in Artificial Intelligence and Logic Programming*, vol. 3, *Nonmonotonic Reasoning and Uncertain Reasoning*, Oxford: Oxford University Press. (Definitive and detailed survey articles on a variety of non-monotonic logics.)

- Horty, J.** (1994) 'Moral Dilemmas and Non-Monotonic Logic', *Journal of Philosophical Logic* 23: 35-65. (An application of non-monotonic logic to problems in moral reasoning.)
- McCarthy, J.** (1959) 'Programs with Common Sense', in D.V. Blake and A.M. Uttley (eds) *Proceedings of the Symposium on Mechanization of Thought Processes*, London: H.M. Stationery Office; expanded version repr. in R. Brachman and H. Levesque (eds) *Readings in Knowledge Representation*, San Mateo, CA: Morgan-Kaufmann, 1985. (An early statement of McCarthy's views.)
- McCarthy, J.** (1980) 'Circumscription - A Form of Non-Monotonic Reasoning', *Artificial Intelligence* 13: 27-39.
- McCarthy, J. and Hayes, P.** (1969) 'Some Philosophical Problems from the Standpoint of Artificial Intelligence', in B. Meltzer and D. Michie (eds) *Machine Intelligence*, vol. 4, Edinburgh: Edinburgh University Press.
- Reiter, R.** (1980) 'A Logic for Default Reasoning', *Artificial Intelligence* 13: 81-132.
- Thomason, R.** (1988) *Philosophical Logic and Artificial Intelligence*, Dordrecht: Kluwer. (Some essays at the intersection of philosophical logic and artificial intelligence.)
- Thomason, R.** (1991) 'Logicism, Artificial Intelligence, and Common Sense: John McCarthy's Program in Philosophical Perspective', in V. Lifschitz (ed.) *Artificial Intelligence and Mathematical Theory of Computation*, San Diego, CA: Academic Press. (An article relating the project of formalizing common-sense knowledge to the logicist programme in philosophy.)

Complexity, computational

The theory of computational complexity is concerned with estimating the resources a computer needs to solve a given problem. The basic resources are time (number of steps executed) and space (amount of memory used). There are problems in logic, algebra and combinatorial games that are solvable in principle by a computer, but computationally intractable because the resources required by relatively small instances are practically infeasible.

The theory of $\mathcal{NP}$ -completeness concerns a common type of problem in which a solution is easy to check but may be hard to find. Such problems belong to the class $\mathcal{NP}$; the hardest ones of this type are the $\mathcal{NP}$ -complete problems. The problem of determining whether a formula of propositional logic is satisfiable or not is $\mathcal{NP}$ -complete. The class of problems with feasible solutions is commonly identified with the class $\mathcal{P}$ of problems solvable in polynomial time. Assuming this identification, the conjecture that some $\mathcal{NP}$ problems require infeasibly long times for their solution is equivalent to the conjecture that $\mathcal{P} \neq \mathcal{NP}$. Although the conjecture remains open, it is widely believed that $\mathcal{NP}$ -complete problems are computationally intractable.

1 Time and space in computation

The theory of complexity analyses the computational resources necessary to solve a problem. The most important of these resources are *time* (number of steps in a computation) and *space* (memory capacity of the computer). This entry is mainly concerned with the complexity of decision problems having infinitely many instances. There is another approach to complexity applicable to individual finite objects, in which the complexity of an object is measured by the size of the shortest programme that produces a description of it (see [Computability and information](#)).

The model for computation chosen here is the Turing machine (see [Turing machines](#)). The *time* for a computation is the number of steps taken before the machine halts; the *space* is the number of cells of the tape visited by the reading head during the computation. Several other models of sequential computation have been proposed. The time and space complexity of a problem clearly depend on the machine model adopted. However, the basic concepts of complexity theory defined in this entry are robust in the sense that they coincide for any reasonable model of sequential computation.

Let Σ be a finite alphabet, and Σ^* the set of all finite strings in this alphabet. A subset of Σ^* is said to be a 'problem'. The size $|s|$ of a string s is its length, that is, the number of occurrences of symbols in it. A function f defined on Σ^* and having strings as its values is computed by a Turing machine M if for any string s in Σ^* , if M is started with s on its tape, then it halts with $f(s)$ on its tape. A problem L is solvable if there is a Turing machine that computes the characteristic function of L (the function f such that $f(s) = 1$ if s is in L and $f(s) = 0$ otherwise). For example, the problem Satisfiability of determining whether a formula of propositional logic is satisfiable or not is solvable by the familiar method of truth tables.

Solvable problems can be classified according to the time and space required for their solution. If f is a computable function, then we say that f is computable in time $T(n)$ if there is a Turing machine M computing f and a constant c such that for any input s , M halts with output $f(s)$ in at most $c \cdot T(|s|)$ steps. A function f is computable in space $S(n)$ if there is a machine M computing f and a constant c such that for any input s , M halts after visiting at most $c \cdot S(|s|)$ squares on its tape. A problem L is solvable in time $T(n)$ if the characteristic function of L is computable in time $T(n)$; L is solvable in space $S(n)$ if the characteristic function of L is computable in space $S(n)$. For example, the truth table method shows that Satisfiability can be solved in time 2^n and space n (we need only enough tape space to evaluate the truth table one row at a time).

One of the most significant complexity classes is the class $\mathcal{P}$ of problems solvable in polynomial time. A function is polynomial-time computable if there exists a k for which f is computable in time n^k . A problem is solvable in polynomial time if its characteristic function is polynomial-time computable. The importance of this class rests on the widely accepted working hypothesis that the class of practically feasible algorithms can be identified with those algorithms that operate in polynomial time. Similarly, the class $\mathcal{PSPACE}$ contains those problems solvable in polynomial space. The class $\mathcal{EXPTIME}$ consists of the problems solvable in exponential time; a problem is

solvable in exponential time if there is a k for which it is solvable in time 2^{n^k} . The class $\mathcal{EXPSPACE}$ contains those problems solvable in exponential space. The problem Satisfiability is in $\mathcal{EXPTIME}$; whether it is in $\mathcal{P}$ is a major open problem.

2 Hierarchies and reductions

The complexity classes defined above represent only a few levels within detailed hierarchies defined by space and time bounds. To state results on the hierarchies, we define a class of functions that includes all the complexity bounds considered here. A function $S(n)$ is said to be fully space constructible if there is a Turing machine that uses exactly $S(n)$ tape cells for any input of length n . With this definition we can state the space hierarchy theorem (Hartmanis, Lewis and Stearns 1965):

If $S_1(n)$ and $S_2(n)$ are fully space constructible functions and if

$$\liminf_{n \rightarrow \infty} \frac{S_1(n)}{S_2(n)} = 0,$$

then there exists a problem solvable in space $S_2(n)$ but not in space $S_1(n)$.

In informal terms, the theorem says that if $S_1(n)$ and $S_2(n)$ are ‘reasonable’ mathematical functions and asymptotically $S_2(n)$ grows faster than $S_1(n)$, then there is a problem solvable in space $S_2(n)$ but not in $S_1(n)$. For example, there is a problem solvable in space n^3 , but not in space n^2 . A similar but slightly weaker time hierarchy theorem can be proved for complexity classes defined by time bounds.

The problems in these theorems that separate the complexity classes are defined by diagonalization. Applications of the theorems to naturally occurring problems in logic and mathematics are obtained by using the tool of efficient reductions. These are the complexity-theoretic analogue of the notion of Turing reducibility in recursion theory.

Let L_1 and L_2 be problems expressed in alphabets Σ_1 and Σ_2 , respectively. L_1 is said to be polynomial-time reducible to L_2 (briefly, reducible to L_2) if there is a polynomial-time computable function f from Σ_1^* to Σ_2^* such that for any s in Σ_1^* , $s \in L_1$ if and only if $f(s) \in L_2$. Many other notions of reducibility have been considered. The important property of polynomial-time reducibility is that it is compatible with the main complexity classes $\mathcal{C}$ considered in this entry; this means that if L_2 is in a complexity class $\mathcal{C}$ and L_1 is reducible to L_2 , then L_1 is also in $\mathcal{C}$.

3 Provably intractable problems

If $\mathcal{C}$ is a complexity class and L is a problem in $\mathcal{C}$ such that any problem in $\mathcal{C}$ is reducible to L , then L is said to be $\mathcal{C}$ -complete. Such problems are the hardest problems in $\mathcal{C}$; if any problem in $\mathcal{C}$ is computationally intractable, then a $\mathcal{C}$ -complete problem is intractable.

The time hierarchy theorem implies that there are problems in $\mathcal{EXPTIME}$ requiring exponential time for their solution, no matter what algorithm is employed. It follows that the same is true for any problem that is $\mathcal{EXPTIME}$ -complete. Such problems can be shown to exist in the area of combinatorial games. For example, we can define generalized chess to be a game similar to standard chess, but played on an $n \times n$ board. The problem of determining of an arbitrary position in generalized chess whether or not the first player has a winning move is $\mathcal{EXPTIME}$ -complete (Fraenkel and Lichtenstein 1981). If M is a Turing machine that solves this problem, there is a constant $c > 1$ such that infinitely many $n \times n$ chess positions exist for which M requires time greater than c^n to determine whether the first player has a win or not. In this sense, chess can be said to be a provably difficult game.

Problems that are $\mathcal{EXPSPACE}$ -complete are also inherently intractable. In algebra, the word problem for commutative semigroups has been proved $\mathcal{EXPSPACE}$ -complete (Mayr and Meyer 1982). An instance of this problem is a finite sequence of equations of the form $s = t$ where the terms s and t are constructed from a set of constants using a single binary operation; the problem to be solved is whether or not the last equation in the sequence can be inferred from the earlier equations by the usual rules for equations, assuming the commutative and associative laws for the operation. It follows that any Turing machine solving this problem requires an exponential amount of space for infinitely many inputs.

Logical theories are a source of solvable decision problems for which no feasible algorithm exists. An example is first-order arithmetic, restricted to addition. The language of this theory is a first-order language based on identity, the addition symbol and the constants 0 and 1. The decision problem consists in determining whether or not a sentence in this language is true in the natural numbers. Presburger in 1930 showed that this problem is decidable by the method of eliminating quantifiers. In 1973 Rabin and Fischer showed that the complexity of the problem is doubly exponential. This means that for any machine solving this problem, there is a constant $c > 0$ such that for infinitely many sentences of length n the machine takes at least $2^{2^{cn}}$ steps to determine whether it is true or not.

Still more powerful lower bounds can be proved by allowing second-order quantifiers. The weakest such extension allows quantifiers ranging over finite sets. Let WS1S be the weak monadic second-order theory of one successor. This theory is formulated in a second-order language with equality, based on the constant 0 and the successor function. The decision problem consists of deciding whether a sentence of WS1S is true in the natural numbers when the second-order quantifiers range over finite sets of natural numbers. It was shown decidable by Büchi in 1960. Define a function $F(n, m)$ by the recursion

$$F(n, 1) = 2^n$$

$$F(n, m + 1) = 2^{F(n, m)}.$$

If d is a real number greater than 0, then $f(d, n) = F(n, [dn])$ represents exponentiation by a stack of $[dn]$ 2's, where $[x]$ is the integer part of x . Albert Meyer in 1972 showed that an algorithm deciding WS1S must use space $f(d, n)$ infinitely often for some $d > 0$. This result shows that the decidability of WS1S is of a purely theoretical nature, so that the theory may be considered undecidable from a practical point of view.

The conclusion of the previous paragraph could be challenged by pointing out that Meyer's lower bound is an asymptotic result that does not rule out a practical decision procedure for sentences of practically feasible size. However, a further result shows that astronomical lower bounds can be proved for WS1S even if we restrict the length of sentences. A Boolean network or circuit is an acyclic directed graph in which each node is labelled with a variable of one of the logical operators 'and', 'not'. The nodes labelled with a variable have no incoming edges, while those labelled 'and' have two, and those labelled 'not' have one. There is exactly one node (the output node) with no outgoing edge. A circuit computes a Boolean function (truth-function) of the variables labelling its input nodes in an obvious way. Meyer and Stockmeyer showed that any such network that decides the truth of all sentences of WS1S of length 616 or less (encoded as binary strings) must contain at least 10^{123} nodes. Even if the nodes were the size of a proton and connected by infinitely thin wires, the network would densely fill the known universe.

4 $\mathcal{NP}$ and the polynomial-time hierarchy

A very common type of practical computational problem takes the form of searching for a solution to a set of conditions, where a proposed solution can be checked easily for correctness. Problems of this kind appear naturally in fields such as operations research, automated theorem proving, computer algebra, linear programming, graph theory and many other areas of discrete mathematics. The features common to such problems have been formalized in the theory of $\mathcal{NP}$ -completeness.

A k -place relation holding between finite strings of symbols in a finite alphabet Σ is said to be polynomial-time computable (see §1 above) if its characteristic function is polynomial-time computable. (The characteristic function of a relation R is the function f such that $f(\langle s_1, \dots, s_k \rangle) = 1$ if $R(s_1, \dots, s_k)$, and $f(\langle s_1, \dots, s_k \rangle) = 0$ otherwise.) Let L be a problem expressed in an alphabet Σ . L belongs to the class $\mathcal{NP}$ if there is a polynomial-time computable relation R and a polynomial $p(x)$ such that for any s in Σ^* , s is in L if and only if there exists a y such that $|y| \leq p(|s|)$ and $R(s, y)$.

The abbreviation ' $\mathcal{NP}$ ' stands for 'non-deterministic polynomial-time', and is derived from an alternative definition of the class using non-deterministic Turing machines. The instructions for such a machine give for each pair consisting of a scanned symbol and the current state a finite set of permissible actions. A language is in $\mathcal{NP}$ if and only if it is accepted by a non-deterministic Turing machine the length of whose computations is bounded by a polynomial in the size of the input. (A string is accepted by such a machine if there is a computation consisting of permissible actions at each step that ends in an accepting configuration.)

Informally, we could describe the relation between $\mathcal{P}$ and $\mathcal{NP}$ as follows. The class $\mathcal{P}$ consists of problems with a feasible decision method; the class $\mathcal{NP}$ consists of problems where we can guess a solution, and check in a feasible amount of time if our guess was correct.

The problem Satisfiability belongs to $\mathcal{NP}$ since we can take the relation $R(s, y)$ to be the relation ‘ y is a truth-value assignment that satisfies the formula s ’. Satisfiability occupies a central position in the theory of $\mathcal{NP}$ -completeness as perhaps the simplest and most natural example of an $\mathcal{NP}$ -complete problem. Cook’s theorem of 1971 showing that Satisfiability is $\mathcal{NP}$ -complete led to the identification of hundreds of $\mathcal{NP}$ -complete problems in computer science and discrete mathematics. The book by Garey and Johnson (1979) includes a partial list of such problems.

If a problem is $\mathcal{NP}$ -complete, then a polynomial-time algorithm for the problem would lead to a polynomial-time algorithm for any problem in $\mathcal{NP}$ so that the equation $\mathcal{P} = \mathcal{NP}$ would be true. A polynomial-time algorithm for the satisfiability problem for propositional logic would lead to polynomial-time algorithms for hundreds of important problems in discrete mathematics; this follows immediately from the definition of $\mathcal{NP}$ -completeness. Because such algorithms have been sought in vain for several decades, it is commonly conjectured that none exists, although no proof of this conjecture has been found. The problem ‘ $\mathcal{P} = \mathcal{NP}$?’ is widely considered the most important open problem in theoretical computer science.

The recursive sets can be identified with the computable sets; if we accept the identification of ‘feasibly computable’ with ‘polynomial-time computable’ then $\mathcal{P}$ is the class of problems with feasible solutions. The recursively enumerable sets are obtained from recursive relations by unbounded quantification; similarly, the problems in $\mathcal{NP}$ are obtained from relations in $\mathcal{P}$ by polynomially bounded quantification. Thus the class $\mathcal{NP}$ is the complexity-theoretic analogue of the class of recursively enumerable sets, and the $\mathcal{NP}$ -complete problems correspond to problems of the highest degree of unsolvability among recursively enumerable sets, such as the halting problem (see [Computability theory §5](#); [Turing machines §3](#)).

The analogy with recursion theory has been extended by defining a complexity-theoretic analogue of the Kleene arithmetical hierarchy of recursion theory. The class Σ_k^p is defined as the class of problems expressible in the form

$$\{s: (\exists y_1, |y_1| \leq p(|s|))(\forall y_2, |y_2| \leq p(|s|)) \dots R(s, y_1, \dots, y_k)\},$$

where there are k alternating bounded quantifiers in the prefix of the defining formula, $p(x)$ is a polynomial and R is a polynomial-time computable relation. In particular, $\mathcal{NP} = \Sigma_1^p$. The class Π_k^p is given by a definition dual to that for Σ_k^p , where bounded existential quantifiers are replaced by bounded universal quantifiers, and universal by existential.

It is conjectured that all the classes in the polynomial hierarchy are distinct, but no proof is known. If $\mathcal{P} = \mathcal{NP}$, then all the classes in the hierarchy collapse to $\mathcal{P}$. All of the classes in the polynomial hierarchy are contained in $\mathcal{PSPACE}$. The equality $\mathcal{P} = \mathcal{PSPACE}$ is consistent with current knowledge, even though $\mathcal{PSPACE}$ appears to be a much more extensive class than $\mathcal{P}$.

The space and hierarchy theorems above were proved by diagonal arguments similar to those used to prove Church’s theorem (see [Church’s theorem and the decision problem](#)). There are reasons to think that diagonalization cannot be used to separate $\mathcal{P}$ and $\mathcal{NP}$ or other low-level complexity classes (the analogy with recursion theory breaks down at this point). These reasons are based on results about oracle machines. If A is a set of strings then a Turing machine with oracle A is defined to be a Turing machine that in addition to the normal type of instruction also has instructions of the form (q_i, σ, q_j, q_k) , where q_i, q_j, q_k are machine states and σ is a symbol from the tape alphabet (see [Turing machines §3](#)). The instruction is interpreted by the machine as follows: when the current state is q_i and the string of symbols to the right of the reading head (starting with the current symbol σ) is in A , then the machine’s next state is q_j , otherwise its next state is q_k . Thus a Turing machine with oracle A has the ability to answer in a single step questions of the type ‘Is this string in the set A ?’ Time and space of a computation by an oracle machine is computed just as for an ordinary Turing machine, counting the time taken for the answer to the oracle query as one step (the oracle answers any query instantaneously).

If A is any set of strings, then by imitating the definitions of the complexity classes above, but substituting ‘Turing

machine with oracle A everywhere for ‘Turing machine’ we can define relativized complexity classes $\mathcal{P}^A$, $\mathcal{NP}^A$ and so on. Then we have the following theorem (Baker, Gill and Solovay 1975):

There is a recursive oracle A for which $\mathcal{P}^A = \mathcal{NP}^A$ and a recursive oracle B for which $\mathcal{P}^B \neq \mathcal{NP}^B$.

The significance of this theorem lies in the fact that known techniques of diagonalization, such as are used in recursion theory, relativize in the presence of oracles. Thus it provides evidence that standard diagonal techniques are inadequate to settle such questions as ‘ $\mathcal{P} = \mathcal{NP}$?’.

See also: [Computability theory](#); [Logical and mathematical terms, glossary of](#)

ALASDAIR URQUHART

References and further reading

- Baker, T., Gill, J. and Solovay, R.** (1975) ‘Relativizations of the $\mathcal{P} = ? \mathcal{NP}$ Question’, *SIAM Journal of Computation* 4: 431-42. (This paper shows that diagonalization techniques very likely do not suffice to answer the $\mathcal{P} = ? \mathcal{NP}$ question.)
- Cook, S.A.** (1971) ‘The Complexity of Theorem-Proving Procedures’, in *Proceedings of the 3rd Annual ACM Symposium on Theory of Computing*, New York: Association for Computing Machinery, 151-8. (The fundamental paper introducing the theory of $\mathcal{NP}$ -completeness.)
- Fraenkel, A.S. and Lichtenstein, D.** (1981) ‘Computing a Perfect Strategy for $n \times n$ Chess Requires Time Exponential in n ’, *Journal of Combinatorial Theory*, series A, 31: 199-213. (This paper shows generalized chess to be a hard game in the computational sense.)
- Garey, M.R. and Johnson, D.S.** (1979) *Computers and Intractability: A Guide to the Theory of $\mathcal{NP}$ -Completeness*, San Francisco, CA: Freeman. (Includes an extensive list of $\mathcal{NP}$ -complete problems from many areas.)
- Hartmanis, J., Lewis, P.M., II and Stearns, R.E.** (1965) ‘Classification of Computations by Time and Memory Requirements’, in *Proceedings of the IFIP Congress 1965*, New York: Spartan Books, 31-5. (The first appearance of the space and time hierarchy theorems.)
- Hopcroft, J.E. and Ullman, J.D.** (1979) *Introduction to Automata Theory, Languages and Computation*, Reading, MA: Addison-Wesley. (A clearly written textbook giving the basic definitions and results of complexity theory.)
- Leeuwen, J. van** (ed.) (1990) *Handbook of Theoretical Computer Science*, vol. A, *Algorithms and Complexity*, Amsterdam: Elsevier. (A collection of detailed survey articles covering many topics not included in this entry, such as parallel complexity and cryptography.)
- Mayr, E.W. and Meyer, A.R.** (1982) ‘The Complexity of the Word Problem for Commutative Semigroups and Polynomial Ideals’, *Advances in Mathematics* 46: 305-29. (One of the earliest papers showing the existence of computationally intractable problems in classical algebra.)
- Sipser, M.** (1997) *Introduction to the Theory of Computation*, Boston, MA: PWS. (An elegant introduction to the basics of computational complexity theory, written by a major figure in the area.)
- Stockmeyer, L.** (1987) ‘Classifying the Computational Complexity of Problems’, *Journal of Symbolic Logic* 52: 1-43. (A very informative survey article.)

Compositionality

A language is compositional if the meaning of each of its complex expressions (for example, 'black dog') is determined entirely by the meanings its parts ('black', 'dog') and its syntax. Principles of compositionality provide precise statements of this idea. A compositional semantics for a language is a (finite) theory which explains how semantically important properties such as truth-conditions are determined by the meanings of parts and syntax. Supposing English to have a compositional semantics helps explain how finite creatures like ourselves have the ability to understand English's infinitely many sentences. Whether human languages are in fact compositional, however, is quite controversial.

It is often supposed that meaning, within a context of use, determines truth-value. Meaning and context are also generally thought to determine what terms refer to, and what predicates are true of - that is, to determine extensions. Because of this, compositionality is generally taken to license a principle of substitutivity: if expressions have the same meaning, and substituting one for another in a sentence does not change the sentence's syntax, the substitution can have no effect on truth. Given this, the idea that meaning is compositionally determined constrains what can be identified with meaning. For example, expressions with the same extension cannot always be substituted for one another without change of truth-value (see [Propositional attitude statements §2](#); [Modal logic, philosophical issues in §3](#)).

Natural languages obey a principle of compositionality (PC) only if something more 'fine-grained' than extensions plays the role of meanings, so that expressions with the same extension can still have different meanings. Some say that no matter how fine-grained we make meaning, there will be counterexamples to the claim that meaning is compositionally determined. For example, 'woodchuck' and 'groundhog' are synonymous. But the sentences 'Mae thinks Woody is a woodchuck' and 'Mae thinks Woody is a groundhog' apparently may diverge in truth-value. (Certainly Mae may assert, with understanding, that Woody is a woodchuck, while denying, with understanding, that Woody is a groundhog. For she can understand 'woodchuck' and 'groundhog' but not realize that they are synonymous.) Such arguments lead to the conclusion that meaning is not compositionally determined (or that different forms cannot share meanings).

There is no consensus about such arguments. All agree PCs need *some* restrictions - they do not, for instance, apply to words in quotational constructions. So if 'thinks' is implicitly quotational, the argument loses its force (see [Use/mention distinction and quotation](#)). If 'thinks', as Frege held, produces a (more or less) systematic shift in the semantic properties of the expressions in its scope, this perhaps reduces the threat to compositionality (see [Sense and reference §§4, 5](#)). Some find such accounts of 'thinks' incredible ([Davidson 1984](#)).

Linguists often understand the claim that a language is compositional as asserting an extremely tight correspondence between its syntax and semantics (see [Syntax](#)). A (simplified) version of such a claim is that (after disambiguating simple word forms), there is, for each (simple) word, a meaning and, for each syntactic rule used in sentence construction, an operation on meanings, such that the meaning of any sentence is mechanically determined by applying the operations on meanings (given by the rules used in constructing the sentence) to the meanings of the simple parts. (Often a host of extra restrictions are incorporated. For example: the operations may be limited to applying function to argument; the order in which operations are applied may be settled by the structure of the sentence.) Some see such principles as providing significant constraints on semantical theories, constraints which may help us decide between theories which are in other respects equivalent (for example, [Montague 1974](#)).

Such strong PCs imply that every ambiguous sentence is either syntactically ambiguous or contains simple expressions which are themselves ambiguous (see [Ambiguity](#)). There are putative counterexamples to this. For instance, it has been alleged that 'The women lifted the piano' has two meanings, a 'collective' one (the group lifted it) and a 'distributive' one (each individual lifted it) (see [Pelletier 1994](#)).

Context sensitivity makes the formulation of PCs a delicate matter. In some important sense, differing uses of 'that is dead' have like meanings and syntax. But obviously they may differ in truth; presumably such differences may arise even within one context. How can this be reconciled with compositionality?

We might (implausibly) say that demonstrative word forms, strictly speaking, lack meanings (only tokens have such). An alternative (see [Kaplan 1989](#)) sees the meaning of the word type ‘that’ as ‘incomplete’, needing contextual supplementation to determine a meaning of the sort other terms have. A third strategy insists that if ‘that is dead but that is not’ is true, its ‘that’s are different demonstrative word types. (Advocates of this view are good at seeing very small subscripts.) This view makes it generally opaque to speakers whether tokens are tokens of the same word type, and thus opaque whether their arguments are logically valid. A fourth strategy sees context as providing an assignment of referents to *uses* of a type in a context. On most ways of working this out, it seems there is no true fleshing out of the principle: the meaning of a sentence type (a) is determined by the meanings of parts and syntax, and (b) determines, in context, the sentence’s truth-value. This approach also seems to undermine the idea that an argument like *that is dead; so that is dead* is formally valid (see [Demonstratives and indexicals](#)).

Even if relatively tight PCs should prove untenable (as claims about natural languages), it would not follow that natural languages did not have compositional semantics - finite theories which assign truth-conditions and other semantically relevant properties to sentence types or their uses. It has been argued that only if they do would it be possible for finite creatures like ourselves to learn them (since the languages involve infinite pairings of sounds and meanings); analogous arguments hold that our ability to understand (in principle) natural languages requires the existence of such semantics. Such arguments apparently presuppose that learning and understanding a language requires knowing a theory from which semantic facts (such as the fact that, for any speaker *x* and time *t*, a use of ‘J’existe encore’ by *x* at *t* says that, at *t*, *x* still exists) are deducible. While such a view is not implausible, alternative plausible views of competence do not support this sort of argument. For example, one might identify linguistic competence with the possession of *syntactic* knowledge (or even just with the possession of syntactic abilities) by someone with appropriate social and environmental relations and behavioural dispositions. Whether natural languages have compositional semantics, and whether the meanings of their sentences are determined simply by the meanings of their parts and syntax, is still not settled.

See also: [Meaning and truth](#); [Semantics](#)

MARK RICHARD

References and further reading

- Davidson, D.** (1984) *Inquiries into Truth and Understanding*, Oxford: Oxford University Press. (Essay 1 argues that learnable languages have compositional semantics; essay 7 argues against Fregean accounts of ‘thinks’.)
- Grandy, R.** (1990) ‘Understanding and the Principle of Compositionality’, in J.E. Tomberlin (ed.) *Philosophical Perspectives*, Atascadero, CA: Ridgeview Publishing Company, vol. 4, 557-72. (Defends compositional semantics against the attacks of Schiffer and others.)
- Kaplan, D.** (1989) ‘Demonstratives’, in J. Almog, J. Perry and H. Wettstein (eds) *Themes From Kaplan*, Oxford: Oxford University Press. (Gives Kaplan’s views of demonstrative meaning.)
- Montague, R.M.** (1974) *Formal Philosophy*, New Haven, CT: Yale University Press. (Essays 6-8 give formulations of compositionality principles. Very difficult.)
- Partee, B.** (1984) ‘Compositionality’, in F. Landman and F. Veltman (eds) *Varieties of Formal Semantics*, Dordrecht: Foris, 281-311. (Evaluates Montague-style principles of compositionality. Difficult but illuminating.)
- Pelletier, F.J.** (1994) ‘The Principle of Semantic Compositionality’, *Topoi* 13: 11-24. (A good introduction to arguments for and against principles of compositionality.)
- Schiffer, S.** (1987) *Remnants of Meaning*, Cambridge, MA: MIT Press. (Contains arguments against the existence of compositional semantics for natural languages.)

Computability and information

The standard definition of randomness as considered in probability theory and used, for example, in quantum mechanics, allows one to speak of a process (such as tossing a coin, or measuring the diagonal polarization of a horizontally polarized photon) as being random. It does not allow one to call a particular outcome (or string of outcomes, or sequence of outcomes) 'random', except in an intuitive, heuristic sense. Information-theoretic complexity makes this possible. An algorithmically random string is one which cannot be produced by a description significantly shorter than itself; an algorithmically random sequence is one whose initial finite segments are almost random strings.

Gödel's incompleteness theorem states that every axiomatizable theory which is sufficiently rich and sound is incomplete. Chaitin's information-theoretic version of Gödel's theorem goes a step further, revealing the reason for incompleteness: a set of axioms of complexity N cannot yield a theorem that asserts that a specific object is of complexity substantially greater than N . This suggests that incompleteness is not only natural, but pervasive; it can no longer be ignored by everyday mathematics. It also provides a theoretical support for a quasi-empirical and pragmatic attitude to the foundations of mathematics.

Information-theoretic complexity is also relevant to physics and biology. For physics it is convenient to reformulate it as the size of the shortest message specifying a microstate, uniquely up to the assumed resolution. In this way we get a rigorous, entropy-like measure of disorder of an individual, microscopic, definite state of a physical system. The regulatory genes of a developing embryo can be ultimately conceived as a program for constructing an organism. The information contained by this biochemical computer program can be measured by information-theoretic complexity.

1 The paradox of randomness

Suppose that persons A and B each give us a string of 32 bits, saying that they were obtained from independent coin flips. If A gives the string

$x = 01101000100110101101100110100101$

and B gives the string

$y = 00000000000000000000000000000000,$

then we will believe A and will not believe B : the string x seems random, but the string y does not. Why? The strings are extremely different from the point of view of regularity. The string y has a maximum regularity which allows us to express it in a very compact way - 'only zeros' - while x appears to have no shortened definition at all. The distinction between regular and irregular strings becomes sharper as the strings get longer: it is clearly easier to specify the number

$10^{10^{10}}$

than the first 1,000 digits of π . Classical probability theory is not sensitive to the above distinction, as strings are all equally probable. However, even Laplace (1749-1827) was, in a sense, aware of this paradox; he suggested that non-random strings are strings possessing some kind of regularity, and since the number of all those strings (of a given length) is small, the occurrence of such a string is extraordinary.

Borel (1909), von Mises ([1928] 1919), Ville (1939) and Church (1940) all elaborated on this idea, but a formal model of irregularity was not found until the mid-1960s. Information-theoretic complexity was introduced independently by Solomonoff (1964), Kolmogorov (1965) and Chaitin (1966). The basic idea is to measure the informational complexity of an object by the size of the smallest program computing it. There are two principal applications: (1) to provide a new conceptual foundation for probability theory and statistics by making it possible to define rigorously the notions of random string and random sequence; and (2) to describe the major limitative theorems in mathematical logic in an information-theoretic form.

Let $A = \{a_1, a_2, \dots, a_Q\}$, $Q \geq 2$, be a finite alphabet, and consider the set of all strings $x_1 x_2 \dots x_n$ of elements in A . We denote by $|x|$ the length of the string x .

It is convenient to think of a computer as a decoding device at the receiving end of a communication channel. Its programs can be viewed as code words, or strings over the alphabet A , and the result of the computation as the decoded message. It is natural to require that successive programs (for example, subroutines) sent across the channel can be separated. This can be achieved by stipulating that programs are ‘self-delimiting’, that is, no meaningful program is a prefix of another. Self-delimiting programs carry within them information about their size, in the sense that they allow the computer to decide when to stop reading the input; they act as ‘atomic’ programs having no significant sub-programmatic structure. Formally, a ‘computer’ is a self-delimiting partial recursive function C which carries strings (programs) into strings (see [Computability theory §§2-3](#)). The (Chaitin) complexity $H_C(x)$ of a string x induced by the computer C is the length of the shortest program for C that outputs x , and halts:

$$H_C(x) = \min\{|y| \mid C(y) = x\}.$$

This notion of complexity is machine-dependent. It is possible to make it ‘asymptotically’ independent of the choice of the computer. The price paid amounts to at most an $O(1)$ uncertainty in the numerical values, and is expressed by the following ‘invariance theorem’: there exists a computer U such that for all computers C one has

$$H_U(x) \leq H_C(x) + O(1).$$

($O(f)$ denotes a function $g(n)$ whose absolute value is bounded by a constant times $f(n)$, for sufficiently large n .)

A universal self-delimiting Turing machine (see [Computability theory §4](#); [Turing machines §2](#)) is a possible candidate for U . Indeed, U simulates the computation of every computer C in the most efficient way: there exists a prefix δ such that the program δx makes U simulate the computation of C on x . The prefix δ depends upon U and C ; it is independent of x . From now on we fix a computer U (called universal) and put $H_U = H$.

The set of meaningful programs for U is an instantaneous code, in the sense that each program can be decoded as soon as we come to the end of the string corresponding to it. In view of the invariance theorem it follows that the set of meaningful programs for U is an asymptotically optimal code for the set of all strings: every string can be coded by such a program, with minimum expected length.

In this context the joint complexity $H(x, y)$ of two strings x, y , defined to be the size of the smallest program that makes U output both of them, satisfies the classical inequality:

$$H(x, y) \leq H(x) + H(y) + O(1),$$

that is, the complexity is asymptotically sub-additive. In the original formulation of these definitions the sub-additivity failed to be true. Chaitin (1974) and Levin (1974) independently found the solution: computers have to be self-delimiting.

2 Random strings

Randomness is not absolute, but relative. It makes more sense to ask ‘To what extent is x random?’ than ‘What is a random string?’.

To define random strings we try to capture the idea that a string is random if it cannot be algorithmically compressed. Formally, a string is ‘random’ if it has ‘almost’ maximal complexity within the set of all strings of the same length. The maximal value of the complexity of strings of length n is given by the formula

$$\max_{|x|=n} H(x) = n + H(\text{string}(n)) + O(1),$$

where $\text{string}(n)$ is the n th string according to the lexicographic order. (Strings are enumerated in order of increasing length: all strings of length one, followed by all strings of length two, and so on.)

We are naturally motivated to adopt the following definition. A string x is (Chaitin) ‘ m -random’ (m a natural number) if

$$H(x) \geq |x| + H(\text{string}(|x|)) - m.$$

x is random if it is 0-random.

To validate the above model of randomness one has to show that random strings defined in this way possess (almost) all properties intuitively associated with randomness. We discuss here only two such properties: (1) the absence of any algorithmic rule for defining random strings; and (2) the equi-distribution of letters and blocks of letters. Property (1) holds true in a strong sense: the set of random strings is ‘immune’, that is, it is infinite and contains no infinite recursively enumerable subset (see [Computability theory §5](#)). The intuition motivating (2) is that in a ‘truly random’ string each letter has to appear with approximately the same frequency, namely Q^{-1} . Moreover, the same property should extend to substrings. Indeed, one proves that all ‘reasonably long’ substrings of a random string have approximately the same frequency, which depends only on the length of substrings.

3 Random sequences

An infinite sequence $\mathbf{x} = x_1x_2 \dots x_n \dots$ is uniquely determined by its prefixes

$$\mathbf{x}(n) = x_1x_2 \dots x_n, \quad n = 1, 2, \dots$$

Is it possible to define the randomness of a sequence by means of the degree of randomness of its prefixes? The answer is affirmative. A sequence $\mathbf{x}$ is random if and only if there exists a natural number c such that $H(\mathbf{x}(n)) \geq n - c$ for all natural $n \geq 1$. This condition is equivalent to the following, apparently stronger, requirement ([Chaitin 1974](#)): a sequence $\mathbf{x}$ is random if and only if

$$\lim_{n \rightarrow \infty} H(\mathbf{x}(n)) - n = \infty.$$

Martin-Löf (1966) made use of constructive probability theory in order to define random sequences. The set of all sequences can be equipped with the product probability (induced by the uniform distribution of letters in the alphabet). A constructive null set (of sequences) is a set which can be recursively covered by recursively enumerable sets having arbitrarily small probabilities. The union of all constructive null sets is still a null set, in fact, a maximal one. It can be shown that the set of random sequences coincides exactly with the complement of the maximal null set. It follows that with probability one every sequence is random. However, from a topological point of view, the set of random sequences is of first Baire category, which is a rather small set.

Random sequences share many properties associated with randomness; for instance, they are Borel normal (every string x appears in a random sequence with the probability $Q^{-|x|}$, which depends only on the length of the string), they pass almost all stochasticity tests, they are incompatible with the slightest form of computability (the possibility to compute an infinite part of a sequence makes that sequence non-random). Random sequences generate, in a sense, all sequences; that is, every sequence is effectively reducible, via a fixed process, to a random one. Randomness is an invariant of number representations: the usual representations of a real number are either random for all bases, or not so for any base.

4 Chaitin’s Omega number

Consider a universal self-delimiting program which, instead of being given a specific program at the beginning of the computation, is fed with a ‘random string’ of bits. Whenever the universal program requests another input bit we just toss a fair coin and input 1 or 0, according to whether the coin comes up heads or tails. Finally we ask the question ‘When the above procedure is begun, what is the probability that the universal program will eventually halt?’. The answer is Chaitin’s Omega number, Ω .

The sequence of digits of Ω is random. This means that we will never know more than a handful of digits of Ω . This number appears to be cabalistic: it can be known of, but not itself known.

The fascination of Ω comes from the fact that it is computationally equivalent to the halting problem (see [Computability theory §5](#)), but contains the same information in a fantastically compressed form. An oracle supplying a few digits of Ω would be of invaluable help in answering many mathematical questions, such as Goldbach’s conjecture (we may construct a computer which halts if and only if it finds an even number greater than 2 which is not the sum of two primes), or Riemann’s hypothesis.

From a practical point of view, even having access to all digits of Ω would not make the task of solving problems very easy, as the time (see [Complexity, computational](#)) to decompress Ω grows extremely rapidly, that is, non-recursively.

The classical opinion that an axiom should be self-evident appears to be too restrictive. Neither Riemann's hypothesis, nor Schrödinger's equation are self-evident, but both are very useful. The possibility of finding new axioms for the positive integers should perhaps get more attention. This quasi-empirical attitude seems to be consistent with Gödel's own views:

axioms need not be evident in themselves, but rather their justification lies (exactly as in physics) in the fact that they make it possible for these 'sense perceptions' to be deduced...I think that...this view has been largely justified by subsequent developments, and it is to be expected that it will be still more so in the future.
([1944] 1990: 121)

5 Randomness in physics

All science is founded on the assumption that the physical universe is ordered and rational. The most powerful expression of this state of affairs is found in the successful application of mathematics to make predictions expressed by means of the laws of physics. Where do these laws come from? Why do they operate universally and unfailingly? Nobody seems to have reasonable answers to these questions. The most we can do is to explain that the hypothesis of order is supported by our daily observations: the rhythm of day and night; the regular ticking of clocks. However, there is a limit to this perceived order. How are we to reconcile these seemingly random processes with the supposed order? How can the same physical process (for example, the spin of a roulette wheel) obey two contradictory laws, the laws of chance and the laws of physics?

As our direct information refers to finite experiments, it is not out of the question to discover local rules, functioning on large, but finite scales, even if the global behaviour of the process is truly random. But, to perceive this global randomness we have to have access to infinity, which is not physically possible. It is important to notice that, consistently with our common experience, facing global randomness does not imply the impossibility of making predictions. Space scientists can pinpoint and predict planetary locations and velocities 'well enough' to plan missions years in advance. Astronomers can predict solar or lunar eclipses centuries before their occurrence. However, we have to be aware that all these results - superb as they may be - are only true within a certain degree of precision, which can hardly be evaluated.

This suggests the idea that the universe might be random (see Calude and Salomaa 1994). Here is an example illustrating this idea. Consider reversible transformations of a string $x, x_{t+1} = T(x_t)$. This equation can be regarded as a deterministic law of motion generating the phase-space trajectory of a simple, idealized system. The complexity $H(x_t)$ tends to increase in the course of evolution, provided the initial state is not too complex. It approaches the equilibrium value consistently with the Boltzmann estimate of the entropy, and fluctuates around the equilibrium. This shows how a reversible, deterministic, dynamic evolution can lead to a significant increase of complexity (see Zurek 1989).

Any physical system may be perceived as a computational process. One could even go further and speculate (with Wheeler, Fredkin, Toffoli) that the universe is a gigantic information-processing system, in which the output is as yet undetermined. In Wheeler's compact form: 'It from bit!', that is, 'it' - every force, particle, and so on - is ultimately present through 'bits' of information. Under this interpretation Ω - which represents the halting probability of the system - acquires physical significance. This constant is not only non-computable, but reflects, in a sense, the indeterministic, chaotic nature of the universe, consistently with Svozil's thesis (1993) according to which chaos in physics corresponds to randomness in mathematics.

6 Logical depth

Charles Bennett's thesis (1988) is that a structure is 'deep' if it is superficially random but subtly redundant; in other words, if almost all its algorithmic probability is contributed by slow-running programs. The value of a message is only slightly related to its information content. For example, a typical sequence of coin tosses has high information content, but little message value. According to Bennett, the 'logical depth' of a message can be measured by the amount of computational work required by its originator, which its receiver is saved from having to repeat. Here we include things which are computable in principle, but only at a considerable cost in resources. Tolstoy's *War and Peace*, Leonardo da Vinci's portrait of *Mona Lisa*, or phase-space sheets that evolve from simple coarse-grained cells via the chaotic dynamics of a simple Hamiltonian are objects with large logical depth.

See also: [Logical and mathematical terms, glossary of](#)

CRISTIAN S. CALUDE

References and further reading

- Atiyah, M. et al.** (1994) ‘Responses to “Theoretical Mathematics: Toward a Cultural Synthesis of Mathematics and Theoretical Physics”’, by A. Jaffe and F. Quinn’, *Bulletin of the American Mathematical Society* 30: 178-207. (Enlightening and provocative essays by, among others, M. Atiyah, A. Borel, G. Chaitin, S. Mac Lane, B. Mandelbroit and R. Thom, on the nature of mathematics.)
- Bennett, C.H.** (1988) ‘Logical Depth and Physical Complexity’, in R. Herken (ed.) *The Universal Turing Machine: A Half-Century Survey*, Oxford: Oxford University Press, 227-58. (Expansion of the material of §6.)
- Borel, É.** (1909) ‘Les probabilités dénombrables et leurs applications arithmétiques’ (Enumerable Probabilities and their Arithmetical Applications), *Rendiconti del Circolo Matematico di Palermo* 27: 247-71. (An important paper inaugurating the modern analysis of randomness.)
- Brisson, L. and Meyerstein, F.W.** (1994) *Inventing the Universe*, New York: State University of New York Press. (Some philosophical consequences of Chaitin’s work.)
- Calude, C.** (1994) *Information and Randomness: An Algorithmic Perspective*, Berlin: Springer. (An informative introduction to basic results in algorithmic information theory in a general setting.)
- Calude, C. and Jürgensen, H.** (1994) ‘Randomness as an Invariant for Number Representations’, in H. Maurer, J. Karhumäki and G. Rozenberg (eds) *Results and Trends in Theoretical Computer Science*, Berlin: Springer, 44-66. (A proof of the invariance of randomness with respect to number representations and a discussion of random real numbers.)
- Calude, C., Jürgensen, H. and Zimand, M.** (1994) ‘Is Independence an Exception?’, *Applied Mathematics and Computation* 66: 63-76. (A topological proof showing that the set of all true but unprovable statements is large.)
- Calude, C. and Salomaa, A.** (1994) ‘Algorithmically Coding the Universe’, in G. Rozenberg and A. Salomaa (eds) *Developments in Language Theory*, Singapore: World Scientific, 472-92. (An informal discussion of the randomness hypothesis discussed in §5.)
- Chaitin, G.J.** (1966) ‘On the Length of Programs for Computing Finite Binary Sequences’, *Journal of the Association for Computing Machinery* 13: 547-69. (Chaitin’s first contribution to the subject.)
- Chaitin, G.J.** (1974) ‘A Theory of Program Size Formally Identical to Information Theory’, in *Abstracts of Papers, 1974 IEEE International Symposium on Information Theory*, Indiana, IN: University of Notre Dame Press; repr. in *Journal of the Association for Computing Machinery* 22: 329-40, 1975. (The first systematic presentation of self-delimiting computers and their complexities; discusses complexity-theoretic characterization of random sequences.)
- Chaitin, G.J.** (1987a) *Algorithmic Information Theory*, Cambridge: Cambridge University Press. (The classic introduction to the subject.)
- Chaitin, G.J.** (1987b) *Information, Randomness and Incompleteness: Papers on Algorithmic Information Theory*, Singapore: World Scientific; 2nd edn, 1990. (An anthology of Chaitin’s seminal papers.)
- Chaitin, G.J.** (1992) *Information-Theoretic Incompleteness*, Singapore: World Scientific. (An original and systematic examination of the incompleteness phenomenon from an information-theoretic point of view.)
- Church, A.** (1940) ‘On the Concept of a Random Sequence’, *Bulletin of the American Mathematical Society* 46: 130-5. (Develops von Mises’ ideas in the framework of computability theory.)
- Gács, P.** (1992) ‘Every Sequence is Reducible to a Random One’, *Information and Control* 70: 186-92. (A proof of the reducibility of any sequence to a random one.)
- Gödel, K.** (1944) ‘Russell’s Mathematical Logic’, in P.A. Schlipp (ed.) *The Philosophy of Bertrand Russell*, Evanston, IL: Northwestern University Press, 123-53; repr. in *Collected Works*, ed. S. Feferman et al., Oxford: Oxford University Press, 1990, vol. 2, 119-41. (Source of the quote in §4.)
- Kolmogorov, A.N.** (1965) ‘Three Approaches for Defining the Concept of “Information Quantity”’, *Problems in Information Transmission* 1: 3-11. (Expounds the considerations leading to the algorithmic definition of random strings.)
- Levin, L.A.** (1974) ‘Randomness Conservation Inequalities: Information and Independence in Mathematical Theories’, *Problems in Information Transmission* 10: 206-10. (Discussion of self-delimiting computers and their complexities.)
- Li, M. and Vitányi, P.M.** (1993) *An Introduction to Kolmogorov Complexity and Its Applications*, Berlin: Springer. (A comprehensive treatise on descriptive complexities.)
- Martin-Löf, P.** (1966) ‘The Definition of Random Sequences’, *Information and Control* 9: 602-19. (An important piece of work describing random sequences in the framework of constructive probability theory.)

- Mises, R. von** (1928) *Wahrscheinlichkeit Statistik und Wahrheit*, Vienna: Springer; trans. *Probability, Statistics and Truth*, London: Allen & Unwin; 2nd revised edn prepared by H. Geiringer, New York: Macmillan, 1961. (Expounds the frequential analysis of randomness, as started by the author in his 1919 paper 'Grundlagen der Wahrscheinlichkeitsrechnung', *Mathematische Zeitschrift* 5: 52-99.)
- Rozenberg, G. and Salomaa, A.** (1994) *Cornerstones of Undecidability*, Englewood Cliffs, NJ: Prentice-Hall. (An enlightening presentation of undecidability and randomness.)
- Schnorr, C.P.** (1981) *Zufälligkeit und Wahrscheinlichkeit. Eine algorithmische Behandlung der Wahrscheinlichkeitstheorie (Randomness and Probability: An Algorithmic Foundation of Probability Theory)*, Berlin: Springer. (The first monograph on algorithmic information theory.)
- Shannon, C.E. and Weaver, C.E.** (1949) *The Mathematical Theory of Communication*, Urbana, IL: University of Illinois Press. (The first and still one of the best books on classical information theory.)
- Solomonoff, R.J.** (1964) 'A Formal Theory of Inductive Inference, Part 1 and Part 2', *Information and Control* 7: 1-22, 224-54. (The first paper dealing with descriptonal complexity.)
- Solovay, R.M.** (1975) 'Draft of a paper (or series of papers) on Chaitin's work...done for the most part during the period of Sept.-Dec. 1974', unpublished manuscript, Yorktown Heights, NY: IBM Thomas J. Watson Research Center. (A highly technical and original presentation of Chaitin's results.)
- Svozil, K.** (1993) *Randomness & Undecidability in Physics*, Singapore: World Scientific. (The first systematic presentation of algorithmic randomness in physics.)
- Ville, J.** (1939) *Étude critique de la notion de collectif (A Critical Study of the Notion of Collective)*, Paris: Gauthier-Villars. (Critical examination of von Mises' model of randomness.)
- Zurek, W.H.** (1989) 'Algorithmic Randomness and Physical Entropy', *Physical Review A* 40: 4,731-51. (A detailed discussion of the relevance of information-theoretic complexity for physics.)

Computability theory

The effective calculability of number-theoretic functions such as addition and multiplication has always been recognized, and for that judgment a rigorous notion of ‘computable function’ is not required. A sharp mathematical concept was defined only in the twentieth century, when issues including the decision problem for predicate logic required a precise delimitation of functions that can be viewed as effectively calculable. Predicate logic emerged from Frege’s fundamental ‘Begriffsschrift’ (1879) as an expressive formal language and was described with mathematical precision by Hilbert in lectures given during the winter of 1917-18. The logical calculus Frege had also developed allowed proofs to proceed as computations in accordance with a fixed set of rules; in principle, according to Gödel, the rules could be applied ‘by someone who knew nothing about mathematics, or by a machine’.

Hilbert grasped the potential of this mechanical aspect and formulated the decision problem for predicate logic as follows: ‘The Entscheidungsproblem [decision problem] is solved if one knows a procedure that permits the decision concerning the validity, respectively, satisfiability of a given logical expression by a finite number of operations.’ Some, for example, von Neumann (1927), believed that the inherent freedom of mathematical thought provided a sufficient reason to expect a negative solution to the problem. But how could a proof of undecidability be given? The unsolvability results of other mathematical problems had always been established relative to a determinate class of admissible operations, for example, the impossibility of doubling the cube relative to ruler and compass constructions. A negative solution to the decision problem obviously required the characterization of ‘effectively calculable functions’.

For two other important issues a characterization of that informal notion was also needed, namely, the general formulation of the incompleteness theorems and the effective unsolvability of mathematical problems (for example, of Hilbert’s tenth problem). The first task of computability theory was thus to answer the question ‘What is a precise notion of “effectively calculable function”?’ Many different answers invariably characterized the same class of number-theoretic functions: the partial recursive ones. Today recursiveness or, equivalently, Turing computability is considered to be the precise mathematical counterpart to ‘effective calculability’. Relative to these notions undecidability results have been established, in particular, the undecidability of the decision problem for predicate logic. The notions are idealized in the sense that no time or space limitations are imposed on the calculations; the concept of ‘feasibility’ is crucial in computer science when trying to capture the subclass of recursive functions whose values can actually be determined.

1 Primitive recursive functions

Let us start with the class of ‘primitive recursive’ functions (from n -tuples of natural numbers to the natural numbers). A precise definition of this class was given by Gödel in 1931 (see van Heijenoort 1967), although he then called it simply the class of recursive functions. The main definitional schema of primitive recursion used in the generation of the elements of this class was well-known in mathematics, having been presented most clearly by Dedekind (1888), who had given a set-theoretic reconstruction that turned out to be of importance also in computability theory. Skolem established in 1923 that most functions in elementary number theory are primitive recursive. The foundational significance of this function class was emphasized by Hilbert and Bernays (1934, 1939): the values of the functions (for any argument) can be determined in finitely many steps, proceeding purely ‘mechanically’. This point is also expressed by saying that the functions are ‘effectively calculable’.

The class $\mathcal{PR}$ of primitive recursive functions is specified inductively and contains as its initial functions the zero-function Z , the successor function S , and the projection functions P_i^n for each n and each i with $1 \leq i \leq n$. These functions satisfy the equations $Z(x) = 0$, $S(x) = x'$ and $P_i^n(x_1, \dots, x_n) = x_i$, for all $x, x_1, \dots, x_n$; x' is the ‘successor’ of x . The class is, first of all, closed under the ‘schema of composition’: given an m -place function ψ and n -place functions $\varphi_1, \dots, \varphi_m$ all in $\mathcal{PR}$, the function ϕ defined by

$$\phi(x_1, \dots, x_n) = \psi(\varphi_1(x_1, \dots, x_n), \dots, \varphi_m(x_1, \dots, x_n))$$

is also in $\mathcal{PR}$; ϕ is then said to be obtained ‘by composition’ from ψ and $\varphi_1, \dots, \varphi_m$. $\mathcal{PR}$ is also closed under the ‘schema of primitive recursion’: given an n -place function ψ and an $(n+2)$ -place function φ both in $\mathcal{PR}$,

the function ϕ defined by

$$\begin{aligned}\phi(x_1, \dots, x_n, 0) &= \psi(x_1, \dots, x_n) \\ \phi(x_1, \dots, x_n, y') &= \varphi(x_1, \dots, x_n, y, \phi(x_1, \dots, x_n, y))\end{aligned}$$

is also a function in $\mathcal{PR}$; ϕ is said to be obtained ‘by primitive recursion’ from ψ and φ . A function is ‘primitive recursive’ if and only if it can be obtained from the initial functions by finitely many applications of the composition and primitive recursion schemas.

The primitive recursive functions do not exhaust the class of effectively calculable functions; that was shown by Ackermann and independently by Gabriel Sudan, who constructed functions that are obviously effectively calculable, but not primitive recursive (see [van Heijenoort 1967](#) for detailed references). The ‘Ackermann function’ can be defined by the following recursion equations.

$$\begin{aligned}\phi_0(x, y) &= S(y) \\ \phi_{n'}(x, 0) &= \begin{cases} x & \text{if } n = 0 \\ 0 & \text{if } n = 1 \\ 1 & \text{otherwise} \end{cases} \\ \phi_{n'}(x, y') &= \phi_n(x, \phi_{n'}(x, y))\end{aligned}$$

Notice that ϕ_1 is addition, ϕ_2 is multiplication, ϕ_3 is exponentiation, and so on; that is, the next function is always obtained by iterating the previous one. For each n , the function $\phi_n(x, x)$ is primitive recursive, but $\phi_x(x, x)$ is not!

We can construct another example of an effectively calculable, but not primitive recursive function ψ by effectively listing all primitive recursive unary functions $\varphi_0, \varphi_1, \varphi_2, \varphi_3, \dots$ and then setting $\psi(x) = 1 + \varphi_x(x)$. Since ψ differs from every φ_i , ψ cannot be primitive recursive; but the effectiveness of our listing of the functions φ_i ensures that ψ can be effectively calculated.

For working with this class of functions it is indispensable to establish further closure properties. To allow, for example, directly general explicit definitions (without the cumbersome detour of projections) of functions such as

$$f(x, y, z) = (xy + z)^2 + 789,$$

one establishes that variables can be permuted, substitutions can be carried out, and that the constant functions $\text{con}_m^n(x_1, \dots, x_n) = m$ are primitive recursive. Beyond addition, multiplication and exponentiation it is useful to know that the following functions are also in $\mathcal{PR}$.

$$\begin{aligned}x! & \quad (\text{factorial}) \\ \text{pred}(x) & \quad (\text{predecessor}) \\ x \dot{-} y & \quad (\text{arithmetic subtraction, } \max(0, x - y)) \\ \overline{\text{sg}}(x) & \quad (= 1 \dot{-} x) \\ \text{sg}(x) & \quad (= 1 \dot{-} \overline{\text{sg}}(x))\end{aligned}$$

Then it is very easy to define the bounded sum of an $(n + 1)$ -place function ϕ in $\mathcal{PR}$ as follows.

$$\begin{aligned}\phi(x_1, \dots, x_n, 0) &= 0 \\ \sum_{x < y'} \phi(x_1, \dots, x_n, x) &= \sum_{x < y} \phi(x_1, \dots, x_n, x) \\ &\quad + \phi(x_1, \dots, x_n, y).\end{aligned}$$

The bounded product is defined similarly.

The functions we listed are useful for the investigation of primitive recursive relations; an n -place relation R between natural numbers is called primitive recursive if and only if its characteristic function χ_R is in $\mathcal{PR}$. (For recursion-theoretic purposes it is convenient to consider 0 to be the truth-value T and 1 the truth-value F.) It is now possible to show directly that the primitive recursive relations are closed under Boolean operations - particularly, ‘

$\neg$ (not), ‘Čamp;’ (and), ‘ $\vee$ ’ (or), ‘ $\rightarrow$ ’ (if...then) - and bounded quantification, that is, ‘ $(\forall x < y)$ ’ (for all x less than y) and ‘ $(\exists x < y)$ ’ (for some x less than y).

Finally, we discuss bounded minimization: if for some x less than y $R(x_1, \dots, x_n, x)$, then the function $\mu x < y. R(x_1, \dots, x_n, x)$ yields the smallest such x , otherwise the function yields y . Define first

$$\phi_1(x_1, \dots, x_n, x) = \begin{cases} 0 & \text{if } (\exists z \leq x) R(x_1, \dots, x_n, z) \\ 1 & \text{otherwise} \end{cases}$$

and notice that

$$\mu x < y. R(x_1, \dots, x_n, x) = \sum_{x < y} \phi_1(x_1, \dots, x_n, x).$$

Thus, if the relation R is primitive recursive, bounded minimization leads to a function in $\mathcal{PR}$.

Given this list of functions in $\mathcal{PR}$ and the closure conditions for primitive recursive relations, we have a convenient framework in which we can show the primitive recursiveness of number-theoretic relations; for example, the characteristic function $\chi_<$ of the ‘less than’ relation ‘ $<$ ’ is given by

$$\chi_<(x, y) = \overline{\text{sg}}(y - x);$$

equality, being a divisor of, and being prime are all defined easily. This indicates that the class is mathematically rather rich.

2 Finite computations

In 1931 and in his Princeton lectures of 1934, Gödel used primitive recursive functions and relations to describe the syntax of particular ‘formal’ theories - after Gödel-numbering the syntactic configurations that make up the theory. Since he strove to arrive at a general concept of formality through the underlying concept of calculability for functions, there was no reason to focus attention on theories whose syntax could be presented primitive recursively. He viewed primitive recursive definability of formulas and proofs as a precise condition which *in practice* sufficed to describe formal systems, but he was searching for a condition that would suffice *in principle*.

In his Princeton lectures, Gödel considered it a very important property that the values of primitive recursive functions can be calculated by a finite computation for arbitrary arguments. He added in a footnote:

The converse seems to be true if, besides recursions according to the scheme (2) [of primitive recursion], recursions of other forms...are admitted. This cannot be proved, since the notion of finite computation is not defined, but it can serve as a heuristic principle.

(1934)

Parenthetically, we note that this remark does not contain a formulation of Church’s thesis (see [Church’s thesis](#)); that (prima facie plausible) reading was explicitly ruled out by Gödel in a letter to Martin Davis (1982). It does, however, indicate the motivation for the considerations in the last part of Gödel’s lectures: there he used quite general forms of recursion to introduce ‘general recursive functions’. (The relation of Gödel’s proposal to suggestions of Herbrand is discussed in Sieg (1994).) These functions are obtained as unique solutions of a system E of equations, and their values must be computable in an equational calculus with just two obviously mechanical rules: the first rule allows the substitution of numerals for variables in any equation derived from E ; the second rule allows the replacement of terms $t(\nu_1, \dots, \nu_n)$ in a derived equation by μ , in case $t(\nu_1, \dots, \nu_n) = \mu$ is a derived equation. (Lower-case Greek letters stand here for numerals.)

This class of general recursive functions very quickly turned out to be characterizable in a variety of ways: Church and Kleene showed the equivalence to λ -definability, Kleene to μ -recursiveness (discussed below), and Turing to computability by his machines (see [Lambda calculus](#); [Combinatory logic](#); [Turing machines](#)). The class introduced by Gödel was used by Church in his first formulation of ‘Church’s thesis’ (see [Sieg 1997](#)). The early, pre-Turing attempts to argue for the thesis are captured in a mathematically concise way through the concept of a function ‘reckonable according to rules’ (*regelrecht auswertbare Funktion*). These functions were introduced by Hilbert and Bernays in the second volume of their *Grundlagen der Mathematik* (1934, 1939); they were characterized as being calculable in deductive formalisms satisfying general recursiveness conditions; the critical

condition required the proof predicate of the formalisms to be primitive recursive. It was shown that the calculations could be carried out in a particular subsystem of arithmetic; Gödel took this fact in 1946 as the basis for his claim that computability is an ‘absolute’ concept (see [Sieg 1994](#)). The analysis of Hilbert and Bernays also clearly revealed the ‘stumbling block’ all these analyses encountered: they tried to characterize the elementary nature of steps in calculations, but could not do so without referring to recursiveness (Church), primitive recursiveness (Hilbert and Bernays), or to very specific rules (Gödel).

Only Turing was able, in his ground-breaking 1936 paper ‘On Computable Numbers’ (see [Davis 1965](#)), to circumvent the stumbling block by focusing on the calculations of a human ‘computer’ proceeding mechanically (see [Turing machines §1](#)). He formulated general boundedness and locality conditions for such a computer and showed that any number-theoretic function whose values can be calculated by a computer satisfying these conditions can actually be computed by a Turing machine (see [Mundici and Sieg 1995](#)). The latter fact is sometimes called ‘Turing’s theorem’. The restricted formulation of Turing machines allows a uniform and simple description of computations; its adequacy for linear computations is guaranteed by Turing’s theorem. The starting point of Turing’s analysis was, however, the mechanical behaviour of a human computer operating on finite configurations in the plane. This behaviour can be described generally and mathematically precisely. To arrive at such a more general description, we take a preliminary step and replace the states of mind by what Turing described as ‘physical and definite counterparts’. This is done by considering ‘state of mind’ not as a property of the working computer, but rather as part of the configuration which is operated on. Turing discussed this replacement very vividly in his 1936 paper:

It is always possible for the computer [in our terminology, the computer] to break off from his work, to go away and forget all about it, and later to come back and go on with it. If he does this he must leave a note of instructions (written in some standard form) explaining how the work is to be continued. This note is the counterpart of the ‘state of mind’. We will suppose that the computer works in such a desultory manner that he never does more than one step at a sitting. The note of instructions must enable him to carry out one step and write the next note.

(‘On Computable Numbers’, in [Davis 1965](#))

This was achieved quite beautifully by Post in 1947, and Post’s approach of describing Turing machine computations is used by Davis ([1958](#)). The configurations on which the machine works are instantaneous descriptions (IDs). These are finite sequences in the alphabet of a Turing machine containing exactly one state symbol; the position of the state symbol indicates which symbol is being scanned. A program of a Turing machine can now be viewed as a set of Post production rules operating on (a single symbol of) such IDs. If in this way of describing Turing machines one replaces finite sequences by finite graphs (with a few well-motivated properties) and the simple Post-Turing operations on one symbol at a time by operations on a fixed finite number of distinguished graphs, then one arrives at the notion of a ‘Kolmogorov machine’. This latter notion, or rather a general concept of algorithm, was introduced by Kolmogorov and Uspensky ([1958](#)): Kolmogorov machines compute exactly the Turing computable number-theoretic functions.

Which number-theoretic functions can be computed by Turing machines or other computing devices? Kleene’s analysis of the equational calculus led to the introduction of the ‘regular minimization’ operator and to an inductive characterization of Gödel’s class of general recursive functions. Suppose the $(n + 1)$ -place function ϕ has the property that for every $x_1, \dots, x_n$ there is a y such that $\phi(x_1, \dots, x_n, y) = 0$. Denote by $\mu y. \phi(x_1, \dots, x_n, y) = 0$ the least such y . Under these conditions, the function ψ given by

$$\psi(x_1, \dots, x_n) = \mu y. \phi(x_1, \dots, x_n, y) = 0$$

is said to be obtained from ϕ ‘by regular minimization’. A function is ‘ (μ) -recursive’ if and only if it can be obtained from the initial functions by a finite number of applications of composition, primitive recursion and regular minimization. That is, the class $\mathcal{R}$ of recursive functions is obtained from $\mathcal{PR}$ by closure under regular minimization.

Up to now we have considered total (that is, everywhere defined) functions and operations that do not lead outside the class of total functions. Gödel (after Kleene) emphasized the importance of partial functions defined only for a subset of $\mathbb{N}$ or $\mathbb{N} \times \dots \times \mathbb{N}$. For example, while $\mu y. xy = 0$ is an equivalent definition of the zero function, an

expression such as $\mu y.x + y = 0$ is undefined for each $x > 0$. For partial functions ϕ we liberalize the regularity condition for minimization:

$$u = \mu y.\phi(x_1, \dots, x_n, y) = 0$$

holds if and only if the following two conditions are satisfied:

(1) For any $i = 0, 1, \dots, u - 1$, $\phi(x_1, \dots, x_n, i)$ is defined and different from 0;

$$\phi(x_1, \dots, x_n, u) = 0. (2)$$

Given $x_1, \dots, x_n$, at most one number u can satisfy both (1) and (2). If no such u exists then the n -tuple $(x_1, \dots, x_n)$ is outside the domain of the function

$$\mu y.\phi(x_1, \dots, x_n, y) = 0$$

or, stated differently, this function is undefined at $(x_1, \dots, x_n)$. A function is ‘partial recursive’ if and only if it can be obtained from the initial functions by a finite number of applications of composition, primitive recursion and (unrestricted) minimization. Note that the partial recursive functions can be enumerated by a partial recursive function; that is obviously not possible in the case of total functions: a standard diagonal construction leads immediately to a contradiction. This fact (for partial recursive functions) and its significance are described in §3 and the beginning of §4.

One can explicitly construct Turing machines for the initial functions and for those operations on programs that correspond to composition, primitive recursion and regular, as well as unrestricted, minimization. Thus, these operations do not lead outside the class of Turing computable functions. This immediately yields the following theorem, showing the power of Turing machines: all partial recursive functions are Turing computable.

3 Effective descriptions

To prove the converse of the theorem just stated, we follow Pythagoras’ ontological prescription and use Gödel’s device to assign numbers e and y as ‘codes’ to Turing machines M and finite sequences of IDs. The coding introduced by Gödel for formal theories in 1931 (and adapted for Turing machines in Davis’ book) will do. Once a machine is coded as a number, the code can be supplied as an input to any machine; this adumbrates the metamorphosis of hardware into software, culminating in the stored-program computer. Programming the first computers virtually amounted to inserting wires into plugboards, but von Neumann - influenced by Turing’s treatment of his paper machines - realized that programs can be coded and stored as strings of symbols in the same way as data.

For any effective Gödel-numbering it is easy, though somewhat tedious, to establish that simple syntactic notions and operations concerning machines and their IDs can be represented by number-theoretic, indeed primitive recursive, predicates and functions. For instance, the ternary predicate $T(e, x, y)$ expressing that y is (the code of) a computation of Turing machine (M with code) e for input x is primitive recursive; also, the unary ‘result-extracting’ function U is primitive recursive, where $U(y)$ is either the number on the tape of the last ID of the computation y or, in case y is not the code of a computation, $U(y) = 0$. (Our choice of the default value 0 is only to guarantee that U is a total function.)

Given a Turing computable function ϕ and a machine M with code e computing ϕ , there are two possibilities for any input x :

Case 1. M terminates. Then there is a (first) computation y^* of M for input x , that is, y^* is the smallest number y such that $T(e, x, y)$; the number $U(y^*)$ read on the tape in the final configuration of y^* coincides with the value $\phi(x)$; writing $\mu y.T(e, x, y)$ as an abbreviation for $\mu y.\chi_T(e, x, y) = 0$, we have

$$\phi(x) = U(\mu y.T(e, x, y)). (3)$$

Case 2. M does not terminate. Then for any y , the predicate $T(e, x, y)$ will be false, that is, $\chi_T(e, x, y) = 1$ for all y , and condition (2) in the definition of the unrestricted μ -operator will never hold. Thus, both $\mu y.T(e, x, y)$ and $U(\mu y.T(e, x, y))$ are undefined.

This establishes (3) for both cases, and we have proved ‘Kleene’s normal form theorem’ for unary Turing computable functions:

Let φ be an n -place number-theoretic function that can be computed by a Turing machine with code e ; then for each $x_1, \dots, x_n$ in the domain of φ we have

$$\varphi(x_1, \dots, x_n) = U(\mu y. T(e, x_1, \dots, x_n, y)).$$

Kleene’s theorem holds also, with the same proof, for functions having any finite number of arguments. Kleene established this normal form theorem not for Turing computable functions, but for total functions that can be calculated in Gödel’s equational calculus. He concluded that Gödel’s general recursive functions are recursive, a conclusion we can draw now for Turing computable partial functions: every Turing computable function is partial recursive.

Using the theorem formulated at the very end of §2, it follows that Turing computability for number-theoretic functions is equivalent to partial recursiveness and, indeed, to all the other characterizations discussed in that section. Kleene’s theorem has most interesting additional consequences, to be discussed in the next section. This is done not just to present elegant mathematical results, but to reinforce the conceptual analysis, as these results (are taken by some to) lend additional support to Church’s thesis (see [Church’s thesis](#)).

4 Basic results

Kleene’s normal form theorem has important consequences for the theory of computability. Consider, first of all, the two-place function

$$\psi(e, x) = U(\mu y. T(e, x, y));$$

ψ is partial recursive and provides an enumeration of all unary partial recursive functions. This is ‘Kleene’s enumeration theorem’. This theorem - together with the equivalence of partial recursiveness and Turing computability - guarantees the existence of a ‘universal’ Turing machine. Indeed, consider a Turing machine $M[\psi]$ computing ψ . For any pair of arguments e and x , $M[\psi]$ interprets its first argument as the code of a Turing machine $M[\varphi]$ (computing a unary partial recursive function φ) and its second argument as the input for $M[\varphi]$. Then $M[\psi]$ proceeds to compute $\varphi(x)$ following the program of $M[\varphi]$. Since $M[\varphi]$ was arbitrary, $M[\psi]$ is able to simulate any Turing machine. Similarly, for each $n = 2, 3, 4, \dots$, there is a partial recursive $(n + 1)$ -place function $\psi^{(n)}$ such that every n -place partial recursive function φ can be written as

$$\varphi(x_1, \dots, x_n) = \psi^{(n)}(e, x_1, \dots, x_n)$$

for a suitable index e for φ .

Given a function $\varphi(p_1, \dots, p_m, x_1, \dots, x_n)$ of $m + n$ variables, let us ‘parametrize’ φ , that is, assign values $p_1^*, \dots, p_m^*$ to the first m variables. We then obtain a new n -place function ϕ as follows:

$$\phi(x_1, \dots, x_n) = \varphi(p_1^*, \dots, p_m^*, x_1, \dots, x_n).$$

Intuitively, ϕ is the ‘slice’ of φ with coordinates $(p_1^*, \dots, p_m^*)$. By Kleene’s enumeration theorem we can write, for suitable indices e and e^* :

$$\begin{aligned} \varphi(p_1, \dots, p_m, x_1, \dots, x_n) = \\ \psi^{(m+n)}(e, p_1, \dots, p_m, x_1, \dots, x_n) \end{aligned}$$

and

$$\phi(x_1, \dots, x_n) = \psi^{(n)}(e^*, x_1, \dots, x_n).$$

The following ‘parameter (or S_n^m -) theorem’ tells us that the slicing operation leading from e to e^* can be performed uniformly and effectively:

For each $m, n > 0$ there is a primitive recursive $(n + 1)$ -place function S_n^m such that

$$\begin{aligned}\psi^{(m+n)}(e, p_1, \dots, p_m, x_1, \dots, x_n) \\ = \psi^{(m)}(S_n^m(e, p_1, \dots, p_m), x_1, \dots, x_n).\end{aligned}$$

This theorem was established by Kleene at first for Gödel's general recursive functions; it amounts in that context to replacing effectively the first m variables of the function φ computed from equations with code e by numerals for $p_1, \dots, p_m$. This theorem is used to prove the following central result, also due to Kleene, the 'recursion theorem':

Let ϕ be a partial recursive function with $(n + 1)$ places. Then there is a number e such that for all $x_1, \dots, x_n$

$$\psi^{(n)}(e, x_1, \dots, x_n) = \phi(e, x_1, \dots, x_n).$$

The theorem finds many uses in establishing that implicitly defined functions are actually partial recursive. For example, letting $n = 1$ and writing ψ instead of $\psi^{(1)}$, an application of the recursion theorem to the function $g(y, x) = y$ immediately yields the 'fixed point theorem':

There is a constant e^* such that for all x

$$\psi(e^*, x) = e^*.$$

Intuitively, for every possible input x , the program with number e^* uses x to output a copy of itself - just like a self-replicating organism. More generally, for every recursive function $r(z)$, we obtain by applying the recursion theorem to the function $\psi(r(z), x)$ a number e such that for all x , $\psi(e, x) = \psi(r(e), x)$.

Kleene's enumeration theorem allows us to reformulate properties of partial recursive functions as number-theoretic properties of their codes. A set P of natural numbers is called an 'input-output property' (of unary partial recursive functions) if and only if, whenever e' is in P and e'' codes the same function as e' (in the sense that $\psi(e', x) = \psi(e'', x)$ for all x in their common domain), then e'' is also in P . The following theorem, 'Rice's theorem', states that non-trivial input-output properties are not recursive: if P is an input-output property such that neither it nor its complement is empty, then P is not recursive.

There are important non-trivial input-output properties of programs. For instance, the set of codes of Turing machines computing total functions is one such property; as a consequence of Rice's theorem we know that there is no Turing machine that decides whether or not a given Turing computable function is total. We will discuss additional undecidable problems in the next section; however, undecidability will be established there in quite different ways.

5 Undecidable problems

'Problems' are identified with (characteristic functions of) one-place predicates or, equivalently, with subsets of $\mathbb{N}$. Let K be the problem corresponding to the question 'Does the computation of machine M with code e terminate, if the input is the natural number e ?'. This problem was formulated by Turing and is known as the 'halting problem' (see [Turing machines §3](#)). A diagonal argument (using the existence of a universal machine) settles the undecidability of K without resorting to Rice's theorem.

The halting problem is the progenitor of a multitude of undecidable problems, arising in virtually all fields of mathematics. Consider, for instance, the following problem. Let there be a presentation P of a group G via generators and relations (P looks like a multiplication table for G) and two expressions v and w built up from these generators, using the multiplication operation and its inverse. Do v and w represent the same element of G ? This is known as the 'word problem for groups'. Novikov and, independently, Boone proved that the problem is undecidable.

Markov proved the undecidability of the following problem for four-dimensional manifolds. Consider a suitable combinatorial description of two manifolds. Are they homeomorphic? In the two-dimensional case the problem becomes decidable; it is not known whether the three-dimensional problem is decidable or not.

Yet another logic-free undecidable problem is given by elementary functions, that is, those functions f of a real variable x which can be built up using exponentials, logarithms, n th roots, trigonometric functions and their

inverses by addition, multiplication, and composition. While the derivative operation does not lead outside the realm of elementary functions, the integral of, say, $(\sin x)/x$ is not elementary. Indeed, Richardson proved that the following problem is undecidable: is the integral of an elementary function elementary?

Finally, the tenth problem in Hilbert's list presented at the International Congress of Mathematicians in 1900 asks for an algorithm to decide of a Diophantine equation - that is, a polynomial equation $p(x_1, \dots, x_n) = 0$, with integer coefficients - whether or not it has integer solutions (see Hilbert 1900). Following important work by J. Robinson, Davis and Putnam, Matijasevic proved in 1970 that this problem is undecidable (see Matijasevic 1993).

Let us go back to the halting problem, K . Recalling the discussion in §3, it is clear that the predicate $K(e)$ can be defined by the statement 'There exists a y , such that $T(e, e, y)$ '. K is thus an example of a 'recursively enumerable' (RE) set of natural numbers: it can be defined by prefixing one existential quantifier to a decidable binary predicate. Trying for each n in $\mathbb{N}$ all possible pairs (e, y) such that $e + y = n$ and singling out e whenever $T(e, e, y)$ is true, we can effectively list all members of K . Recursively enumerable sets were named because of the following equivalent characterization: a subset of $\mathbb{N}$ is RE if and only if it is either empty or the range of a (primitive) recursive function. Furthermore, a set R is recursive if and only if both R and its complement are RE. This latter fact allows a different route into computability theory: recursive enumerability is taken as the basic concept, and recursive sets are defined as those RE sets whose complements are also RE; this is done in Post's approach.

The importance of K among RE sets stems from its being 'complete', that is, maximally difficult among all RE sets; by this we mean that every RE set can be effectively reduced to K . A problem P is effectively reducible to Q if and only if there is a recursive function f such that for all natural numbers x , $x \in P$ if and only if $f(x) \in Q$. Thus, if we want to see that x is a solution of problem P we compute $f(x)$ and test whether $f(x)$ solves Q ! Clearly, Q is at least as difficult as P , in the sense that effective solutions for Q yield effective solutions for P . By effectively reducing the halting problem K to the *Entscheidungsproblem* (decision problem) E , Turing concluded from the undecidability of K that also E cannot be decidable. Turing's reduction is nothing but a transcription of the ternary predicate $T(e, e, y)$ into a first-order arithmetical theory with sufficient demonstrative power.

The halting problem K , the most complex among RE problems, is the simplest in an infinite sequence of manifestly undecidable problems, the so-called 'jumps'. In order to obtain the jump hierarchy, the concept of computation is relativized to sets of natural numbers whose membership relations are revealed by 'oracles'. The jump K' of K , for example, is defined as the halting problem, when an oracle for K is available. This hierarchy can be associated in a most informative way with definability questions in the language of arithmetic: all jumps can be defined by increasingly complex arithmetical formulas, and all arithmetically definable sets are reducible to some jump. The above construction underlies the arithmetic hierarchy introduced by Kleene and Mostowski in the fifties.

Certain interesting sets of natural numbers are not definable by arithmetic formulas; one example is the set of all Gödel numbers of true arithmetic statements. If this set were arithmetically definable, one could formulate arithmetically the liar sentence that expresses its own falsity. This observation of Gödel and Tarski (see Gödel's theorems) is the cornerstone for proving the incompleteness of every formal theory of arithmetic that contains symbols for addition and multiplication and has semantically sound axioms and rules of inference: no matter which (true) statements we choose as axioms, and no matter which inference rules we adopt (provided they are truth-preserving) there are statements which, albeit true for the natural numbers, are not provable in the theory. Arithmetic truth can be defined, however, in second-order languages that allow quantification over functions.

6 Physical steps

Turing appealed in his analysis of mechanical calculability to the limitations of the human sensory apparatus; he claimed that the justification for his thesis lies ultimately 'in the fact that the human memory is necessarily limited'. This remark is not expanded on at all, and we can only speculate as to Turing's understanding of this 'fact': did he have in mind more than the spatial limitations for 'encoding' finite configurations? If not, the restrictions can be motivated by physical considerations. The question of whether there are ways of getting around the effect of such physical limitations is discussed with references to the literature by Mundici and Sieg (1995); and Mundici (1981).

Assume that a Kolmogorov machine operates on configurations containing z different symbols, each symbol being physically ‘coded’ by at least one atom - an altogether reasonable assumption. Then there must be at least z pairwise disjoint regions containing the codes (of the symbols). Otherwise, the electron clouds of two different codes might overlap, making the codes indistinguishable. Let c and a denote the speed of light and Bohr’s radius of the hydrogen atom, respectively, so that $a/c = 0.176 \times 10^{-18}$ seconds. It follows that the codes will be contained in a volume of at least $z(4/3)\pi a^3$ cubic metres; that forces the diameter of the machine to be larger than $2az^{\frac{1}{3}}$ metres - the diameter being the largest possible distance between two codes in this volume. Let f be the frequency of our machine; thus, $1/f$ is the time available for each computation step. Since signals cannot travel faster than light and since a computation step involves the whole configuration, it follows that f cannot exceed $c/(2az^{\frac{1}{3}})$ steps per second. Thus, we obtain the inequality

$$fz^{\frac{1}{3}} < 2.828 \times 10^{18} \text{ steps per second,}$$

which points out a fundamental incompatibility between high number of codes (that is, size of configurations) and high computational speed. The operations of Kolmogorov machines are thus restricted in complexity, as they have to lead from distinguished graphs to distinguished graphs; and within the given physical boundaries only finitely many different graphs are realizable.

If we focus on physical devices and analyse machine, not human mechanical, computability we have to take into account the possibility of parallel procedures as incorporated, for example, in cellular automata. Gandy (1980) provided the first conceptual analysis and a general description of parallel algorithms. These algorithms are thought to be carried out by ‘discrete deterministic mechanical devices’, that is, machines satisfying the physical assumptions explicit in our discussion above; in Gandy’s words,

The only physical assumptions made about mechanical devices...are that there is a lower bound on the linear dimensions of every atomic part of the device and that there is an upper bound (the velocity of light) on the speed of propagation of changes.

(1980: 126)

He formulated axiomatic principles for these devices and proved that whatever can be calculated by devices satisfying the principles (‘Gandy machines’) is also computable by a Turing machine.

See also: [Logical and mathematical terms, glossary of](#)

DANIELE MUNDICI
WILFRIED SIEG

References and further reading

- All the classic papers referred to in this entry have been reprinted in either van Heijenoort (1967) or Davis (1965); careful discussions of the history of the subject are given by Gandy in Herken (1988) and by Sieg (1994).
- Cutland, N.** (1980) *Computability - An Introduction to Recursive Function Theory*, Cambridge: Cambridge University Press. (An informative introduction to basic results.)
- Davis, M.** (1958) *Computability and Unsolvability*, New York: McGraw-Hill; repr. New York: Dover, 1982. (A classic, detailed presentation of the basic parts of computability theory.)
- Davis, M.** (ed.) (1965) *The Undecidable: Basic Papers on Undecidable Propositions, Unsolvability Problems and Computable Functions*, Hewlett, NY: Raven Press. (Anthology of the fundamental papers on the subject by Gödel, Church, Turing, Kleene, Rosser and Post.)
- Davis, M.** (1982) ‘Why Gödel Didn’t Have Church’s Thesis’, *Information and Control* 54: 3-24. (Provides fascinating historical information on the developments in Princeton in 1933 to 1936.)
- Dedekind, R.** (1888) *Was sind und was sollen die Zahlen?*, Braunschweig: Vieweg; repr. in *Gesammelte mathematische Werke (Collected Mathematical Works)*, ed. R. Fricke, E. Noether and Ö. Ore, Braunschweig: Vieweg, 1932, vol. 3, 335-90; trans. W.W. Beman (1901), ‘The Nature and Meaning of Numbers’, in *Essays on the Theory of Numbers*, New York: Dover, 1963; trans. *What are Numbers and What Should They Be?*, Orono, ME: Research Institute for Mathematics, 1995. (The ‘modern’ set-theoretic foundation of number theory.)
- Frege, G.** (1879) *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*, Halle: Nebert; trans. ‘*Begriffsschrift, a Formula Language, Modelled Upon That of Arithmetic, for Pure Thought*’, in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge,

- MA: Harvard University Press, 1967, 1-82.(This small booklet includes the most significant step from Aristotelian to modern logic.)
- Gandy, R.** (1980) 'Church's Thesis and Principles for Mechanisms', in J. Barwise, H.J. Keisler and K. Kunen (eds) *The Kleene Symposium*, Amsterdam: North Holland, 123-45.(Axiomatic characterization of machine, not human mechanical, computability including parallel computations.)
- Gödel, K.** (1934) Notes of Princeton Lectures, in *Collected Works*, ed. S. Feferman *et al.*, vol. 1, *Publications 1929-1936*, New York and Oxford: Oxford University Press, 1986.(Quoted in §2. German originals and English translations.)
- Heijenoort, J. van** (ed.) (1967) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879- 1931*, Cambridge, MA: Harvard University Press.(Anthology of fundamental papers in logic starting with Frege's *Begriffsschrift* (1879) and ending with Gödel and Herbrand.)
- Herken, R.** (ed.) (1988) *The Universal Turing Machine: A Half-Century Survey*, Oxford: Oxford University Press. (This book includes informed, systematic, enlightening, historical and provocative essays by, among others, Kleene, Gandy, Davis, Feferman and Penrose.)
- Hilbert, D.** (1900) 'Mathematische Probleme. Vortrag, gehalten auf dem internationalen Mathematiker-Kongress zu Paris 1900', *Nachrichten von der königlichen Gesellschaft der Wissenschaften zu Göttingen. Geschäftliche Mitteilungen* 253-97; repr. in *Archiv der Mathematik und Physik* 1: 44-63, 213-37, 1901; trans. 'Mathematical Problems', *Bulletin of the American Mathematical Society* 8: 437- 79, 1902.(The most influential list of open mathematical problems in the twentieth century.)
- Hilbert, D. and Bernays, P.** (1934, 1939) *Grundlagen der Mathematik (Foundations of Mathematics)*, Berlin: Springer, 2 vols; 2nd edn, 1968, 1970.(A most comprehensive presentation of the work in proof theory done in the Hilbert school and of related developments; for example, volume 2 includes the first detailed proof of Gödel's second incompleteness theorem.)
- Kolmogorov, A.N. and Uspensky, V.A.** (1958) 'On the Definition of an Algorithm', trans. in *American Mathematical Society Translations* 2 (21): 217-45, 1963.(This paper gives a very interesting analysis of 'algorithms'.)
- Matijasevic, Y.V.** (1993) *Hilbert's Tenth Problem*, Cambridge, MA: MIT Press.(A beautiful and thorough presentation of the solution to Hilbert's tenth problem, originally solved by Matijasevic in a paper of 1970, 'Enumerable Sets are Diophantine'.)
- Mundici, D.** (1981) 'Irreversibility, Uncertainty, Relativity and Computer Limitations', *Il Nuovo Cimento, Europhysics Journal* 61 B (2): 297-305.(Discusses physical limitations of computer steps, arising from relativity and quantum mechanical principles, such as Pauli indistinguishability.)
- Mundici, D. and Sieg, W.** (1995) 'Paper Machines', *Philosophia Mathematica* 3: 5-30.(A nontechnical essay on the development of the main ideas of computability, also covering physical computation.)
- Neumann, J. von** (1927) 'Zur Hilbertschen Beweistheorie' (On Hilbert's Proof Theory), *Mathematische Zeitschrift* 26.(This paper includes the central contribution of von Neumann to the development of proof theory: he proves the consistency of a fragment of number theory.)
- Odifreddi, P.** (1992) *Classical Recursion Theory*, Amsterdam: North Holland, 2 vols.(A comprehensive modern treatise on computability theory.)
- Rogers, H.** (1967) *Theory of Recursive Functions and Effective Computability*, New York: McGraw-Hill.(The standard reference on recursion theory before Odifreddi (1992) and Soare (1987).)
- Sieg, W.** (1994) 'Mechanical Procedures and Mathematical Experience', in A. George (ed.) *Mathematics and Mind*, Oxford: Oxford University Press, 71-117.(Examination of the conceptual analyses underlying computability theory, in particular in Turing's fundamental paper; presents the historical development in detail and emphasizes connections to investigations in the foundations of mathematics.)
- Sieg, W.** (1997) 'Step by Recursive Step: Church's Analysis of Effective Calculability', *Bulletin of Symbolic Logic*.(Describes in detail the approach Church took to the conceptual issues underlying his work on computability and undecidability. New historical material is taken into account to complement (and correct) the discussion in Davis (1982).)
- Soare, R.I.** (1987) *Recursively Enumerable Sets and Degrees*, Berlin and New York: Springer.(Brings the reader to the frontiers of research on computable functions, featuring Post's problem, oracles, priority methods and lattices of RE sets.)

Computer science

At first sight, computers would seem to be of minimal philosophical importance; mere symbol manipulators that do the sort of things that we can do anyway, only faster and more conveniently. Nevertheless, computers are being used to illuminate the cognitive abilities of the human and animal mind, explore the organizational principles of life, and open up new approaches to modelling nature. Furthermore, the study of computation has changed our conception of the limits and methodology of scientific knowledge.

Computers have been able to do all this for two reasons. The first is that material computing power (accuracy, storage and speed) permits the development and exploration of models of physical (and mental) systems that combine structural complexity with mathematical intransigence. Through simulation, computational power allows exploration where mathematical analysis falters. The second reason is that a computer is not merely a concrete device, but also can be studied as an abstract object whose rules of operation can be specified with mathematical precision; consequently, its strengths and limitations can be systematically investigated, exploited and appreciated. Herein lies that area of computer science of most interest to philosophers: the theory of computation and algorithms. It is here where we have learned what computers can and cannot do in principle.

1 Turing machines and computability

In his classic paper ‘On Computable Numbers, with an Application to the *Entscheidungsproblem*’ (1936) Alan Turing investigated the problem of characterizing the computable real numbers. The real numbers are those numbers that correspond to the points on a straight line: the positive and negative integers, the fractions or rational numbers, and finally, the irrational numbers, such as π and the square root of 2. Whereas the values of the integers and fractions can be specified using a finite set of symbols and computed exactly, the same is not true for the irrationals. Some of the irrationals like π or the square root of 2 can be computed to as many decimal places as we like; other irrationals, however, cannot be computed at all. Intuitively, the difference between the two kinds of numbers lies in the fact that the former can be determined to arbitrarily high degrees of precision, in Turing’s words, ‘using finite means’, whereas to obtain an arbitrarily precise value of the latter sort of number would require infinite means.

Turing proceeds to give a precise account of what it means to compute ‘using finite means’. Imagine a machine that at any given time is in one of a finite number of states $Q = \{q_0, q_1, \dots, q_i\}$. At the same time, the machine is scanning a square of a finite but extendible tape on which is printed one of a finite number of symbols $S = \{s_0, s_1, \dots, s_k\}$. The machine is capable of doing just three things. First, it writes a symbol on the square (possibly the same symbol that it is scanning). Next, it moves one square to the left or one square to the right. And finally, it enters a (possibly) new state. The key idea is that what the machine does is determined only by its current state and the symbol currently being scanned. Exactly what the machine is to do in a given situation is specified by a rule which is stated in the form of a ‘quintuple’ such as the following:

$q_2 * \# R q_3,$

which may be read, ‘if in state q_2 and scanning the symbol $*$, then write the symbol $\#$, go to the right and enter state q_3 ’. A Turing machine is formally defined as a set of such instructions, subject to the proviso that no two quintuples begin with the same two items. (This ensures that there is no ambiguity as to what the machine is to do when in a certain internal state and scanning a given symbol.) Often, the state q_0 is designated as the starting state of the machine, and the symbol s_0 is considered to be the ‘blank’ symbol B . A Turing machine halts when it is in a certain internal state q_i , scanning a symbol s_i , and there is no quintuple beginning with $q_i s_i$. (What to do, what to do...?) Using these conventions, and assuming that the machine is started on the right-most symbol of a finite string of 1s, then the machine described by the quintuples:

$q_0 1 1 L q_0$

$q_0 B 1 L q_1$

will (as the reader can verify easily) move to the left leaving all of the 1s intact until it finds the first blank (‘ B ’) which it will then replace by a 1, move one more square to the left, and then halt (see [Turing machines](#)).

Let us call the symbols $S^* = \{s_1, s_2, \dots, s\}$ the (proper) *alphabet* of the Turing machine. (Notice that the ‘blank’ symbol q_0 is hereby excluded.) The *words* of this alphabet are now defined as the set of all the finite strings of the members of the alphabet. A function on this alphabet is simply a map that takes any word to a unique word; that is, a map from the set *words* to itself. (For simplicity’s sake, only functions taking a single argument will be considered here.) Such a function is said to be *Turing computable* just in case there is a Turing machine M which, when started on the right-most letter of a word, eventually halts leaving on its tape the string that is the function’s value on the original word. Roughly, a Turing computable function is a function that can be computed by some Turing machine.

Of all the functions on an alphabet, how many are Turing computable? With the aid of Cantor’s theory of infinite sets (see **Cantor’s theorem**), we can answer this question fairly easily. Basically, since a Turing machine is determined by (and can thus be regarded as defined by) a finite set of quintuples, Turing machines can (in principle) be ordered ‘alphabetically’. So there are only a countable infinity of Turing machines and hence at most a countable infinity of Turing computable functions (since each Turing computable function requires at least one Turing machine to compute it). But there are an uncountable infinite number of functions from (any) set of words to itself. Hence, although there are an infinite number of Turing computable functions, there are an uncountable number of functions that are not Turing computable. So to the question, ‘of all the functions from words to words, how many are computable?’, the answer is ‘a countable infinity, that is, almost none of them’. Later, we shall examine an interesting function - the Halting function - that is not computable.

Although we have considered functions from and to the words of any finite alphabet of symbols, the same results may be obtained by confining our attention to functions from the set N of natural numbers $0, 1, 2, \dots$, to itself. To see this, think of the symbols $S^* = \{s_1, s_2, \dots, s\}$ as the digits $\{0, 1, \dots, i-1\}$. Then, corresponding to a finite string of letters in S^* , we have a string of digits, which can then be regarded as representing a natural number using base j . In this way, the words of S^* become mapped to numbers in N , and functions from *words* to *words* become functions from N to N . This allows us to restrict our attention to the Turing computable functions from N to N without any important loss of generality.

2 The Church-Turing thesis

In work carried out prior to Turing’s 1936 paper (presented to the American Mathematical Society in 1935), Alonzo **Church** (1936) had defined a class of computable functions from the set $N = \{0, 1, 2, \dots\}$ of natural numbers to itself which he called the ‘recursive’ functions. The recursive functions are defined, roughly, as those that can be obtained from a few basic computable functions (in the sense that we have algorithms for them) by applying ‘computable’ operations such as functional composition or recursion. For example, one of the basic functions is the *successor* function which takes a number n to $n + 1$. Another is multiplication. Composing these two functions we obtain $(n + 1)m$. Church’s basic functions, Turing showed, were Turing computable, and the operations for their combination could also be mimicked by Turing machines. For example, the function $mn + 1$ can be computed by a Turing machine that chains together the Turing machines that compute mn and $n + 1$. In this way, Turing showed that the recursive functions are all Turing computable. In addition, he also proved the converse: any Turing computable function from N to N is recursive. As a result, the Turing computable functions and the recursive functions are identical (Boolos and Jeffrey 1982, chs 7 and 8).

In his 1936 paper, Church had conjectured that the recursive functions were just the computable functions from and to the natural numbers N . In his paper of the same year, Turing also conjectured that the functions computable by his machines were also just the computable functions. Since the two sets of functions are identical, these conjectures have since been joined and are known as the *Church-Turing thesis*, namely,

Any computable function from N to N is recursive, that is, Turing computable (see also **Church’s thesis**).

Let us see what reasons there are for holding this thesis to be correct.

First of all, notice that in the above statement of the thesis, the first occurrence of the term ‘computable’ has not been given a precise definition. The intuitive idea is that a function is computable if we can calculate its output using means that are of the same sort that we use to do calculations in arithmetic. But what, exactly, are these ‘means’? We have examples galore, but no precise general account of what sort of procedures these examples

embody. We could, of course, simply take Church's recursive functions or Turing's machine to *define* just those procedures. But in that case, the Church- Turing thesis becomes a tautology.

So let us stick with 'computable' as a pre-analytic term and see where that gets us. Consider the following example. Suppose that we consider a new kind of machine model, very much like a Turing machine, except that instead of a one-dimensional tape, the new machine operates on a 'board' of two dimensions, a potentially infinite checkerboard. Such a *2D Turing machine*, as I shall call it, will have quintuples just like a Turing machine, except that the move instructions *R*, *L* are expanded to eight possibilities: *N*, *S*, *E*, *W*, *NE*, *NW*, *SE*, and *SW*. Imagine that we now write out a set of quintuples and proceed to use them in the same way that we previously used a Turing machine - say, to determine what results from putting the string *!#** down on the board and starting at the eastmost symbol (*). Clearly, if we could show that such a machine would stop when provided with any such input, then we would agree that a function was thereby *computed* by this 2D Turing machine. In short, the same intuitions which assure us that any function which is Turing computable is *computable* in the ordinary sense - the uncontested converse of the Church-Turing thesis - also work here to yield the conviction that any function that is 2D Turing machine computable is *computable*. But if this be granted, it can now be seen that there is - at least in principle - a way to show the *falsity* of the Church-Turing thesis: produce a machine model, such as a 2D Turing machine, that clearly *computes*, and then prove that there are some functions that *it* can compute that a standard Turing machine cannot.

In this spirit (and perhaps just for the fun of it), there has been a great deal of work on developing alternative models of computation. The 2D Turing machine model given above, along with its *n*-dimensional counterparts, can be shown to be no more powerful than a standard linear tape machine. Adding finitely many tapes with their own read/write heads does not help either: any function computable by these enhanced machines can be proven to be computable (perhaps more slowly) by a lowly standard Turing machine. Other models that are not simple variations of Turing machines have also been explored, such as abacus or register machines, post systems and generalized digital computers. While these alternative models might usefully focus on one or more features of computation systems, they all are incapable of computing a function that cannot likewise be computed by a 'plain vanilla' Turing machine. As a result, although there is no way of proving the Church-Turing thesis to be correct, it is widely held to be so.

The Turing machines that we have considered so far have all been dedicated to some particular computational task. A *universal Turing machine* (UTM) would be capable of computing anything that a Turing machine can. In his 1936 paper, Turing showed how to construct machines of this type, and, roughly, here is how they work. A section of the tape of a UTM is set aside (perhaps surrounded by *s or some other symbol in its alphabet reserved for this purpose) for the program that the machine is to execute. This program will be another Turing machine's quintuples (probably suitably encoded). A suitable input is now provided in some input area of the tape, and the UTM passes back and forth from the program and input areas to the work area, performing the computation in accord with the program on its tape. When one considers that much the same sort of activity is precisely what undergraduates do when, in a purely mechanical way, they work out on paper the behaviour of a Turing machine provided by the instructor, then the existence of a universal Turing machine should not come as much of a surprise. What is more remarkable is that such a machine is described in some detail in Turing's 1936 paper. Later work has produced some very concise UTMs, such as Marvin Minsky's, which uses only four symbols and seven states (Minsky 1967: §14.8). Turing's basic idea - put the program on the tape (in the temporary memory, RAM) of a universal computer - was later taken up by von Neumann, and became one of the pillars of current digital computer architecture.

3 The Halting problem

As we have already seen, 'most' functions from *N* to *N* are not computable. Let us now look at a particular function that bears this affliction, the Halting function. Given a particular Turing machine or more generally, a particular program, it would be useful to be able to know whether that program will halt when provided with a given input. Of course, in some cases, the answer is obvious. The problem here is to find a general method, another program, that can compute the answer to the halting problem for *any* given input program and its input. Putting the matter in terms of Turing machines: is there a Turing machine that, when provided with a (suitably encoded version of a) program on its tape, together with an input to that program, will output, say, a '1' if the program will

halt on that input and a '0' otherwise. (It is assumed here that, after delivering its verdict, the Halter machine itself then halts.)

It turns out that no such Turing machine exists. This result is sometimes referred to as the undecidability of the Halting problem. One reason this is so important is that it is the basis of showing that a number of other such problems are undecidable (by reducing them to the Halting problem); another reason is that it shows that there is a strong connection between the failure of computability and a wide class of unsolved problems in number theory. For example, in 1742 Goldbach conjectured that every even number (greater than 2) can be expressed as the sum of two prime numbers; to this day, his conjecture has been neither proven nor disproven. Now it is a simple matter to write a program that will check an even number to see if it is expressible as the sum of two primes, and do this for successively greater even numbers. If such a program were to find an even number not equal to the sum of two primes, then it could be designed to halt and print out some message like 'Goldbach has been refuted!'. On the other hand, if Goldbach was right, then the program will go on checking even numbers indefinitely, and thus never halt. If the Halting problem were decidable, then such problems - and there are a great many such problems - could (in principle, at least) be settled according to the Leibnizian saying: 'Let us compute!' (For a survey of results on uncomputable functions, see Harel 1992, ch. 8.)

4 Problems that are computationally intractable

Even when we consider only computable functions, there are some problems whose solutions will probably forever remain beyond our reach. A classic example is the travelling salesperson problem. Imagine a set of N towns laid out on a map, together with the various roads connecting these towns. The length of each road is given. The problem now is to find the shortest itinerary for the salesperson to take which starts at a given town, returns to that town at the finish, and visits every other town exactly once. Clearly, the solution is computable: enumerate each of the possible paths, computing and recording its length along the way, and return as the answer the path requiring the minimum total distance. The problem is that the number of steps required to execute this algorithm goes up drastically as the number of towns involved gets larger. Thus, for 100 towns, the number of steps required by the best current programs is greater than 2^{100} , a very large number. (Assuming our algorithm executes one million steps per second, such an algorithm would require over 10^{16} years to complete its task; the accepted upper limit on the age of the universe is, however, only around 15 billion, that is, 15×10^9 years.)

The travelling salesperson problem is only one of a wide class of problems, called *NP complete* problems, that require, as far as we know, solutions on the order of N steps. (The reason that *NP complete* problems are so-called may be found in Harel 1992, ch. 7.) If any one of the problems in this class could be shown to be solvable by an algorithm that was, say, linear in N or polynomial in N , then the same would be true for all the problems in this class. While it is widely believed that the *NP complete* problems require algorithms that work in exponential time, this has not yet been proven, and remains one of the most difficult unsolved problems in the theory of computation.

Problems having an input of size N that require 2^N or more steps for their solution are said to be *intractable*. While there remains some (dim) hope for the tractability of the *NP complete* problems, there are many other problems that are provably intractable, and thus require algorithms operating in exponential time or worse. For example, a complete examination of the game tree of checkers could determine whether or not the side that moves first has a guaranteed winning strategy. Consider now generalized checkers, played not on an 8×8 board, but on an $N \times N$ board. It has been shown that the problem of determining the existence of a guaranteed first-mover winning strategy is intractable in N ; that is, there is no algorithm for solving this problem which does so in fewer than K^N steps ($K > 1$). This is not the end of it, however. Some problems can be shown to be solvable in at best doubly exponential time 2^{2^N} , or even worse. Similar results have been established for problems that involve unreasonably large amounts of memory space (Harel 1992, ch. 7).

The situation is not altogether bleak. There are tractable algorithms for solving *NP complete* problems that give correct or nearly correct answers most of the time. For example, there are tractable algorithms designed to tackle the travelling salesperson problem that routinely come up with paths that are surely far shorter than most routes. Another possible way around intractability is an algorithm that generates the correct answer (say, to a decision problem) with a certain high probability, regardless of the size N of the input. More precisely, suppose that we can show that if the answer to our question is truly 'yes', then there is an algorithm that will output 'yes' with a

probability $P > K$, regardless of the size N of the input; on the other hand, if the correct answer is ‘no’, then our algorithm will quickly yield that result. Clearly, such an algorithm is almost as good as one that provides answers whose correctness can be guaranteed.

One such probabilistic algorithm was presented by Solovay and Strassen (1977). It is designed to test an arbitrary natural number for compositeness. (Does the number have nontrivial factors?) Number theorists have been able to show that if a number is composite (that is, not prime), then over half of the numbers less than that number will have a quickly computable property W that will be a conclusive witness to that fact. Thus if a number is not composite, no such witness will exist. A candidate N for compositeness is now input into the algorithm and a number less than N is chosen at random. If the randomly chosen number has the ‘witness’ property W , then the original number N must be composite, so the program outputs ‘Composite!’ and halts. On the other hand, if the randomly chosen number does not have the witness property W , then the algorithm chooses another number less than N at random. If 100 or so random choices fail to turn up a witness to N ’s compositeness, then the program outputs ‘Not composite!’. A little thought shows that since over $1/2$ of the numbers less than N will have W if N is composite, then if we fail to find such a witness after 100 random trials, the chances of N ’s being composite are less than $1/2^{100}$, a very small chance indeed. Provided we are willing to take a small risk of error, this method results in a very fast probabilistic algorithm for testing compositeness.

Whether the existence of such an algorithm shows that probabilistic algorithms can tame intractable problems is, however, still unclear. The hitch is that the compositeness problem is not *known* to be intractable. Furthermore, there are no cases of problems *known* to be intractable that have succumbed to a probabilistic algorithm (although some *NP* complete problems have been tamed). The issue of whether or not there are probabilistic algorithms that can solve (provably) intractable problems remains open. It *has* been shown, however, that probabilistic algorithms cannot tame problems that are not computable, a result that follows from the fact that any probabilistic algorithm can be simulated deterministically.

The theory of deterministic algorithms has also contributed to our understanding of what apparently stands in direct opposition to determinism: randomness (see [Randomness §4](#)).

See also: [Complexity, computational](#); [Computability and information](#); [Computability theory](#); [Turing reducibility and Turing degrees](#)

JOHN WINNIE

References and further reading

- Boolos, G.S. and Jeffrey, R.** (1974) *Computability and Logic*, Cambridge: Cambridge University Press; 2nd edn, 1982. (A classic, easy to read text on computation theory and its relation to logic.)
- Chaitin, G.** (1969) ‘On the Length of Programs for Computing Finite Binary Sequences: Statistical Considerations’, *Journal of the Association for Computing Machinery* 16: 145-59. (Chaitin’s improvement on the Kolmogorov definition of algorithmic complexity.)
- Church, A.** (1936) ‘An Unsolvable Problem of Elementary Number Theory’, *American Journal of Mathematics* 58: 345-63; repr. in M. Davis (ed.) *The Undecidable*, New York: Raven Press, 1965, 89-107. (Church’s original paper on the lambda calculus approach to the theory of computation.)
- Davis, M.** (ed.) (1965) *The Undecidable*, New York: Raven Press. (A sourcebook of classic papers on computation theory.)
- Davis, M., Sigal, R. and Weyuker, E.** (1983) *Computability, Complexity, and Languages: Fundamentals of Theoretical Computer Science*, Boston, MA: Academic Press; 2nd edn, 1994. (A handbook, perhaps the handbook, of the current state of computation theory. A work of moderate technical difficulty.)
- Garey, M. and Johnson, D.** (1979) *Computers and Intractability*, New York: W.H. Freeman. (A detailed but only moderately technical account of the theory of computational intractability.)
- Harel, D.** (1992) *Algorithmics*, Reading, MA: Addison-Wesley. (An excellent, easy to read survey of the current state of computation theory.)
- Kolmogorov, A.N.** (1965) ‘Three Approaches to the Quantitative Theory of Information’, *Problems in Information Transmission* 1: 1-7. (The paper in which Kolmogorov introduces algorithmic measures of complexity.)
- Li, M. and Vitanyi, P.** (1993) *An Introduction to Kolmogorov Complexity and its Applications*, New York: Springer. (A thorough, fairly technical, account of Kolmogorov and Chaitin’s ideas on complexity.)

- Minsky, M.** (1967) *Computation: Finite and Infinite Machines*, Englewood Cliffs, NJ: Prentice Hall.(A lively, easy to read survey of computation theory as things stood in the early 1960s.)
- Solomonoff, R.J.** (1964) 'A Formal Theory of Inductive Inference, Part 1 and Part 2', *Information and Control* 7: 1-22, 224-54.(The first of the classic trio of papers (Solomonoff, Kolmogorov, Chaitin) on the foundations of algorithmic information theory.)
- Solovay, R. and Strassen, V.** (1977) 'A Fast Monte-Carlo Test for Primality', *Society for Industrial and Applied Mathematics Journal of Computing* 6: 84-5.(An early probabilistic algorithm for testing a number for primality, discussed in the text.)
- Turing, A.** (1936, 1937) 'On Computable Numbers, with an Application to the *Entscheidungsproblem*', *Proceedings of the London Mathematical Society*, series 2, 42: 230-65; corrections, 43: 544-6; repr. in M. David (ed.) *The Undecidable*, New York: Raven Press, 1966, 116-54.(The classic paper on Turing machines: difficult, brilliant, containing many errors.)

Consequence, conceptions of

The idea of one proposition's following from others - of their implying it - is central to argument. It is, however, an idea that comes with a history attached to it, and those who blithely appeal to an 'intuitive' or 'pre-theoretic' idea of consequence are likely to have got hold of just one strand in a string of diverse theories. This entry introduces the main alternatives - to call them rivals would be too strong, since it suggests that they are necessarily in competition with one another. Simply put, consequence may be conceived as a relation that is or is not modal in character, and is or is not formal. Thus for Aristotle, consequence is both necessary and formal; for Chrysippus it is necessary but not formal; for Bolzano and Tarski it is formal but not necessary; and for Philo and Russell it is neither necessary nor formal. Conceptions of consequence that are neither necessary nor formal are also needed if justice is to be done to deduction in science, the law and daily life. Cutting across all these other differences there is a perennial controversy about relevance. Does implication always require a full-blooded connection between premises and conclusion, or may it hold simply because of some property of either separately, for example, because the premises are impossible?

1 Aristotle

Alone among the persons discussed here, Aristotle never defined his conception of consequence. A definition can be supplied, however, that brings out its distinctive features: necessity, formality and relevance. Namely, a number of premises imply a conclusion if, for all premises $P, \dots$ and conclusions Q of the same logical form as the given ones, if $P, \dots$ are (or were) true, then necessarily Q is (would be) true.

The ingredient of necessity here (unlike formality) features in Aristotle's definition of a syllogism. It is required by his demand that proofs should produce 'understanding' (*epistēmē*), coupled with his claim that understanding something involves seeing that it cannot be otherwise. Hence, a proof needs to establish the necessity as well as the truth of its conclusion, which means not only starting from necessarily true axioms but proceeding by steps that preserve necessity as well as truth.

A part of Aristotle's project of axiomatizing science was to separate out those lines of argument that are applicable to all the sciences because they depend purely on topic-neutral terms such as 'every', 'some' and 'not'. These arguments are 'logical' in the special sense in which Aristotle sometimes - not always approvingly - uses the word to signify 'abstract'. The distinction between logical - that is, topic-neutral - and extralogical terms is not clear cut, but he sidesteps the difficulties by selecting a fixed handful of logical terms for study at any one time. Logicians have done the same ever since, splitting consequence in predicate logic, modal logic and so on, rather than lumping them into one broad relation of logical consequence.

Since an extralogical term does no work in making a logical argument valid, the argument remains valid if the term is replaced by any other. In other words, as the definition emphasizes, logical consequence is 'formal'. This feature is exploited by Aristotle's innovation of inference schemes, in which letters take the place of the non-working terms. The combination of formality and an artful selection of working terms creates opportunities for systematization and a mathematical treatment of the subject.

The definition uses counterfactuals because that seems to give the best fit with Aristotle's position on relevance. Where P is impossible and Q unconnected with it, 'If P were true Q would be true' is quite impossible to assess, and such cases are duly excluded by his dictum that premises only imply conclusions involving the same terms. But if a scheme is valid whenever its instances can be assessed (as with the syllogistic moods), he assumes that it holds universally; that is, even when the premises are impossible.

2 Chrysippus, Philo and Diodorus

Our knowledge of the fourth century BC dialecticians Chrysippus, Philo and Diodorus is fragmentary and second-hand, but they appear to have made two lasting contributions to the subject of consequence, one distinctly unhelpful, the other of enormous importance (see [Philo the Dialectician §2](#); [Diodorus Cronus §3](#)).

Expressions of the form 'follows from' or 'implies' need to be completed by sentences embedded in quotation marks or 'that' clauses. As any logic teacher knows, this is not something that comes naturally, and the simpler 'If

P , Q form is a convenient shorthand. The Greek dialecticians seem to have turned it into a regular convention, making the conditional *the* vehicle for discussing consequence. Apart from the other drawbacks of the convention, its blurring of an object language-metalanguage distinction is liable to make a study of consequence look like a contribution to propositional logic. This is especially unfortunate, since their great contribution was to propose the first definitions of consequence and inaugurate a debate about it.

For Chrysippus (the attribution is not certain but strongly probable), premises imply a conclusion if their truth is incompatible with its falsity. ‘A man is running’ thus implies ‘An animal is moving’. As a class, such analytic implications are logical in a broad sense but not in Aristotle’s special one; they are necessary without being formal.

For Philo of Megara, an implication holds if - for whatever reason - it is not the case that the premises are true and the conclusion false.

Diodorus Cronus’ definition is that ‘it neither was nor is possible’ that the premises be true and the conclusion false. But since he equates the possible with what is or will be true, this amounts to saying that an implication holds if and only if Philo’s condition is satisfied at every time.

Little has survived of the debate itself, alas, but relevance was certainly one issue, for we know that where P is impossible, ‘If P , not- P ’ was acceptable to Diodorus but not Chrysippus. Another issue is likely to have been the extreme weakness of Philo’s definition, to which Diodorus’ looks like a critical rejoinder. To defend (and perhaps motivate) Philo’s definition, one need only ask why there should be a higher standard for the steps of an argument than for its starting point. Premises do not have to be necessarily true, so why should conclusions have to follow *necessarily*? Weak as it is, Philo’s definition captures what *Principia Mathematica* (1910-13) calls the ‘essential property’ of implication; that what is implied by a true proposition is true. It still seems far-fetched to suggest that, for example, all true propositions imply one another, but Philo has been defended even on this point. Russell (1919) distinguishes the presence of an implication from its practical utility for inference. To infer any truths from any others would indeed be *valid* (his word, his italics), but if they are already known to be true there will be no need to infer them, so that inference will in fact only take place if Philo’s condition is known to obtain independently, a much stricter requirement.

3 Medieval logic

Medieval logicians tended to follow the ancient dialecticians’ expository convention. Thus their ‘sequences’ (*consequentiae*) tend to be conditional propositions, though their discussion is appropriate to consequence: a sequence is typically said to be ‘sound’ (*bona*; *valet*) rather than true. The conditional convention lingered into modern times, as witnessed in the ambiguous status of Gentzen’s ‘*Sequenzen*’ and Russell’s ‘material implication’.

As this suggests, medieval logicians were familiar with all the ancient conceptions of consequence. They classified them into ‘formal’, that is, Aristotelian, and ‘material’. Material consequence was subdivided into ‘*simpliciter*’ (absolute), that is, Chrysippian; and ‘*ut nunc*’ (as things are), that is, Philonian, the maxim ‘*ex falso sequitur quodlibet*’ (from a falsehood anything follows) being a reminder of the latter. A few writers created a further subdivision, ‘*per accidens*’ (incidental), to cover a Diodorean conception.

One medieval contribution has left a lasting mark on the subject. We have seen that different ancient conceptions produced different answers about relevance. The medieval development of propositional logic appeared to settle the question by producing a proof that *any* conclusion Q really does follow from an impossible premise P . The proof runs from P to P or Q , and thence, together with the necessary truth not- P , to Q . Ironically, the credit for this is generally given not to, say, William of Soissons, but to C.I. Lewis, a logician and historian of logic who had no time for the Middle Ages. The topic is still controversial: Bolzano argued that nothing at all follows from an impossibility, and not all modern logicians accept the Lewis proof (see [Relevance logic and entailment §1](#)). However, most follow the medieval line, which, when used to reform Chrysippian consequence, produces the familiar textbook formula for validity: ‘It is impossible for the premises to be true and the conclusion false’.

4 Bolzano

We saw that Aristotelian consequence is formal, but this does not depend on its being specifically logical: a

comparable relation based on *any* division between working and non-working terms will automatically be formal with respect to the latter. Bolzano exploits this insight in the context of a theory of propositions that are - unlike the propositions mentioned elsewhere in this entry - non-linguistic items, composed of 'ideas' (1837). A conclusion is said to follow from a number of premises with respect to an arbitrary choice of ideas if every substitution for those ideas that turns the premises into true propositions also turns the conclusion into a true one. The other logical notions are relativized in a similar way. Bolzano is happy to include 'logical consequence' as a special case, that is, consequence when every extralogical idea is treated as substitutable. But his heart is in proving theorems about relative consequence in general. (Since he also broke with the old conditional expository convention, his results are unambiguously intelligible to the modern reader.)

Bolzano's definition is non-modal. Indeed, a modal relation seems to be ruled out by his theory of ideas. His emphasis on the immutability of ideas and propositions and their reference to the actual, suggests that he cannot allow, as Frege can, that a false proposition *P* could be true in different circumstances. In different circumstances *P* itself would stay false; what would be true would, by hypothesis, be a different proposition. (Moreover, his resolutely non-linguistic stance prevents him from identifying it.) It follows that a modal definition of consequence on the lines of 'Whatever the circumstances might be, if *P* were true *Q* would be true' will not work (and cannot be rescued by appealing to 'the proposition corresponding to *P*').

5 Tarski

A century after Bolzano, Tarski took a similar approach to logical consequence, but where Bolzano had worked with non-linguistic propositions, he dealt in sentences. Given this choice, the obvious first step would be to replace Bolzano's substitutions of ideas by substitutions of words. This would produce the following definition (labelled (F) by Tarski, 1936):

(F) A conclusion follows logically from a number of premises if every substitution for the extralogical vocabulary that makes the premises true, makes the conclusion true.

Indeed, Quine takes just this line when he defines a logical truth as a sentence that remains true under every substitution for its extralogical vocabulary. But the problem with such definitions is that in a particular language there may not be enough words of the right kind, so that an inference may come out as logically valid, or a sentence as logically true, simply because the counter-instances happen to be inexpressible. Quine plays down the difficulty by claiming that it will not arise if the language is 'reasonably rich', that is, capable of expressing arithmetic. His claim may be correct for logical truth but, as Boolos (1975) has shown, it does not extend to consequence.

Tarski had already come across this problem in his work on the concept of truth in formalized languages (1933). A first shot at the truth-conditions of a universally quantified sentence would be to say that it is true if and only if every particular instance of it is true; but this will not work if there are insufficient constants available, for example, if not every individual has a name. Tarski's solution was to get round any shortage of genuine constants by treating free variables as roving constants, capable of taking each object of the appropriate type (individuals, relations, and so on) in turn as their 'value'. Using the idea of a sentential function, in which free variables replace constants, he defined the 'satisfaction' (making true) of a sentential function by an assignment of values to its variables. The truth of $(x)Fx$ can then be equated with the universal satisfaction of the sentential function Fx .

It was a natural step to apply the same ideas to the definition (F). For brevity, let a 'model' of a sentence be an assignment of values that satisfies the sentential function obtained by replacing all the extralogical constants in the sentence by variables. Then, Tarski says, a conclusion follows logically from a number of premises if every model of the premises is a model of the conclusion.

This definition leaves open where to draw the line between logical and extralogical. In particular, it leaves one free to treat 'thing' (as in 'everything *Fs*', 'something *Fs*') as a logical constant or not. On the second option each assignment of values involves choosing a 'domain of individuals' as value for the variable that replaces 'thing'. The notion of an assignment of values then becomes that of a structure, in the familiar sense of a collection of objects structured by various relations. If too the logical constants are restricted to a mathematically amenable few (for example, $\supset$ instead of 'if'), the notions of satisfaction and model may be defined in purely set-theoretic terms

(see [Model theory](#)). Because of this, the resulting *model-theoretic* conception of consequence has been enthusiastically adopted by many mathematical logicians. Its intimate identification with set theory does, however, expose it to the troubles of that discipline. For, on the prevailing conception, the universe of sets is too large to be a set, and is therefore ineligible to be the domain of any model. For special reasons, first-order logic is unaffected, but the model-theoretic approach to second-order consequence has a question mark over it.

In keeping with the positivism of his day, Tarski wanted nothing to do with modality. (Granted, he uses ‘must’ when formulating (F) as a scheme, but only to indicate that the scheme holds universally; indeed, he stresses that (F) involves no notion more problematic than that of truth.) The non-modal aspect of his definition has been attacked, by Kneale (1961) and subsequently by Etchemendy (1990). A debate is called for, but it will be more fruitful if it asks for what purposes necessity is an essential ingredient of consequence. For example, someone who does not endorse Aristotle’s doctrine of proof and *epistēmē* may well be content with proofs that establish the bare truth of theorems, and it is not obvious that this requires a modal relation of consequence. Again, one motive for developing (say) group theory in the abstract is to derive laws that can be applied to all possible groups. But here a ‘possible group’ does not mean a non-existent structure composed of possibilities, but an actual structure waiting to be recognized as a group (compare with ‘a possible hiding-place’). Once more, a modal conception of consequence is not obviously required.

6 Enthymemes

For a genuinely pre-theoretic notion of consequence one needs to look outside the logic books, and to remember too that a philosopher’s conception of modality is not the only legitimate one: lawyers, scientists and others deal in *real* possibilities, not mere notional logical possibilities. In such circles, arguments take place on the basis of some propositions and some inferences which are common ground and taken for granted; call these an argument’s ‘basis’. But this is not peculiar to everyday argument. A striking example is the predicate calculus, with its built-in presumption that domains of individuals are non-empty (and the corollary that what it holds out as logical consequence fails the test of formality with respect to the extralogical term ‘thing’). Indeed *all* deduction has to proceed on some basis or other, for if every step were challengeable, argument could never get moving.

Historically, adherents of one conception of consequence have tried to exclude any other by the doctrine of the ‘enthymeme’, or incomplete argument. If an argument ‘*P*, so *Q*’ looks sound but does not measure up to the favoured standard, some part of its basis (call it *B*) is declared to be a missing, ‘suppressed’ premise. The original arguer was either incompetent or deluded: their argument should have been, or really was, ‘*P*, *B*, so *Q*’. This expanded version will satisfy the favoured standard, which is thereby vindicated. As well as using this weapon against everyday consequence, logicians have turned it on each other. Traditional logicians use it to claim that all consequence is really syllogistic. Formal logicians of every stripe use it to claim that all necessary consequence is really formal (that is, someone who takes for granted the transitivity of ‘before’ is guilty of suppressing a premise, though someone who takes for granted the transitivity of ‘implies’ is not).

The self-serving character of the enthymeme doctrine is plain. Perhaps what is needed is, rather, an inclusive theory of consequence that will accommodate all kinds of basis for arguments. Let *B* be any basis; then a number of premises imply a conclusion on the basis of *B* if the conclusion belongs to the closure of the premises with respect to *B*. In the spirit of Bolzano, one can go on to develop the general theory of consequence relative to a basis. It needs to be complemented, however, by a quite different kind of discussion, about what bases are appropriate to what purposes and circumstances. In this connection (setting aside any differences over relevance), the broad conceptions of consequence described in the previous sections can be seen as lumping together a class of eligible bases. Thus Chrysippian consequence covers - more exactly, it is the union of - consequence taken over all bases composed exclusively of analytic propositions and steps. Similarly, Philonian consequence is the union of consequence taken over all bases composed of true propositions and truth-preserving steps, though its paradoxical characteristics suggest that this is not the best way of dealing with so diverse a class.

See also: [Logical and mathematical terms, glossary of](#)

TIMOTHY SMILEY

References and further reading

Bolzano, B. (1837) *Wissenschaftslehre*, trans. R. George, *Theory of Science*, Oxford: Blackwell, 1972. (Discussed

- in §4. The relevant sections of this readable but wide-ranging book are §§12, 19, 48, 66, 125, 147, 154-5, 182.)
- Boolos, G.** (1975) 'On Second-Order Logic', *Journal of Philosophy* 72: 509-27. (The appendix includes the technical result cited in §5. Despite the title, Boolos' result, like Quine's, relates to first-order logic.)
- Carnap, R.** (1934) *Logische Syntax der Sprache*, Vienna: Springer; trans. A. Smeaton, *The Logical Syntax of Language*, London: Kegan Paul, 1937. (Carnap briefly introduces consequence as closure with respect to arbitrary 'P-rules'. Tarski had anticipated the idea of closure under rules but, as his 'On the Concept of Logical Consequence' reveals, he was unwilling to take it beyond effective, finitary rules, that is, unwilling to use it to define consequence rather than deducibility.)
- Etchemendy, J.** (1990) *The Concept of Logical Consequence*, Cambridge, MA: Harvard University Press. (Uses a distinction between 'representational' and 'interpretational' semantics to develop Kneale's criticism of Tarski, from a standpoint of uncritical adherence to Chrysippian consequence, relevance and all.)
- Kneale, W.C.** (1961) 'Universality and Necessity', *British Journal for the Philosophy of Science* 12: 89-102. (Criticizes Bolzano and Tarski.)
- Kneale, W. and Kneale, M.** (1962) *The Development of Logic*, Oxford: Oxford University Press. (Chapters 2-4 include an invaluable introduction to the early history of consequence.)
- Russell, B.A.W.** (1919) *Introduction to Mathematical Philosophy*, London: Allen & Unwin. (Cited in §2. Philonian consequence is defended in chapter 14.)
- Smiley, T.** (1995) 'A Tale of Two Tortoises', *Mind* 104: 725-36. (Expands the theme of §6 of this entry.)
- Tarski, A.** (1933) *Pojęcie prawdy w językach nauk dedukcyjnych*, Warsaw; trans. J.H. Woodger, 'The Concept of Truth in Formalized Languages', in *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*, Oxford: Oxford University Press, 1956, 152-278. (Cited in §5.)
- (1936) 'O pojęciu wynikania logicznego', *Przegląd filozoficzny* 39: 58-68; trans. J.H. Woodger, 'On the Concept of Logical Consequence', in *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*, Oxford: Oxford University Press, 1956, 409-20. (This non-technical article is where Tarski expounds the idea of formal, non-necessary consequence discussed in §5. Although the option of varying the domain of individuals is not mentioned, it is given a thorough airing in his more technical 1933 piece.)
- Whitehead, A.N. and Russell, B.A.W.** (1910) *Principia Mathematica*, Cambridge: Cambridge University Press, vol. 1, §1; 2nd edn, 1925. (Cited in §2.)

Constructivism in mathematics

Constructivism is not a matter of principles: there are no specifically constructive mathematical axioms which all constructivists accept. Even so, it is traditional to view constructivists as insisting, in one way or another, that proofs of crucial existential theorems in mathematics respect constructive existence: that a crucial existential claim which is constructively admissible must afford means for constructing an instance of it which is also admissible. Allegiance to this idea often demands changes in conventional views about mathematical objects, operations and logic, and, hence, demands reworkings of ordinary mathematics along nonclassical lines. Constructive existence may be so interpreted as to require the abrogation of the law of the excluded middle and the adoption of nonstandard laws of constructive logic and mathematics in its place.

There has been great variation in the forms of constructivism, each form distinguished in its interpretation of constructive existence, in its approaches to mathematical ontology and constructive logic, and in the methods chosen to prove theorems, particularly, theorems of real analysis. In the twentieth century, Russian constructivism, new constructivism, Brouwerian intuitionism, finitism and predicativism have been the most influential forms of constructivism.

1 Orientation

A call for ‘constructive existence’, standardly thought a hallmark of constructivism in mathematics, represents a general, revisionary attitude towards certain crucial statements of existence. That attitude is expressible as E:

(E) A mathematical statement $\exists x M(x)$, if crucial, is admissible only if a well-defined a can be constructed such that $M(a)$ is admissible, where a denotes a .

For E to be revisionary, the constructivist must indicate which claims are to be crucial, provide accounts of well-definedness and constructibility, and respond to any disruptions in traditional logic incited by E. It is in terms of these requirements that the outlines of influential forms of constructivism are demarcated.

Well-definedness. E need not govern all existential statements. Forms of constructivism are earmarked by choices of crucial statements, statements to which E is to be applied. Predicativists impose a version of E on claims about classes, but sometimes ignore E in elementary number theory. Russian constructivists put all their mathematics under the aegis of a strict reading of E.

As a call for revision, E is nugatory without strictures on well-definedness and constructibility. Conventionally, whenever $\exists n A(n)$ holds of natural numbers, there is a determinate m which is the least number for which $A(x)$ holds. However $\exists n A(n)$ is proved, one might then think to ‘construct’ an n such that $A(n)$ by using the computable function which outputs the number m on any input. So, unless the constructivist supplies further explanation, any proof of $\exists n A(n)$ might seem to satisfy E.

In general, well-definedness is explained in terms of canonical notations. Some constructivists count a natural number canonically denoted once a tally-mark numeral is available for it. Constructivist E. Bishop identified well-defined integers with number descriptions which are convertible into base-ten numerals. A real number may be deemed well-defined once a finitary recipe for approximating it is at hand. A predicativist about classes refuses classes not admitting definition in suitably predicative terms.

If well-definedness depends upon supplies of notations, it is natural to imagine constructive mathematical realms as nominalistic in that everything in them be nameable. It is also common to picture constructive realms as expanding over time as more notations are created. Some see in this image a contemporary rendering of Aristotle’s notion of potential infinities, of domains which are inexhaustible not because there now exist infinitely many elements in them but because, over time, more elements always appear. Explicit treatment of time enters infrequently into constructive mathematics itself, Brouwer’s use of the creative subject affording a celebrated exception (see [Intuitionism §3](#)).

Constructions. Constructibility is explicated in functional or operational terms. In general, a value is constructible just in case there is a recipe for computing that value in principle. Again, definite restrictions are needed if E is to

fulfil a desire which often motivates constructivism: that acceptable definitions in mathematics lead to algorithmic computation. This desire is explicit in E^* , a generalization of E many constructivists endorse.

(E^*) The statement $\forall x \exists y A(x, y)$ is, if crucial, admissible only when there is a construction o which computes, given any well-defined value a for x , a well-defined $o(a)$ such that $A(a, o(a))$ is admissible.

Conventionally, axioms of choice ensure that, when $\forall x \exists y A(x, y)$ holds, there is a function f such that $\forall x A(x, f(x))$ holds too. Hence, talk of ‘constructions’ must amount to more than talk of conventional functions. Once again, forms of constructivism are differentiated in the kinds of construction allowed. Traditional intuitionists demand that the class of constructions include only those operations given by computing recipes which humans can, in principle, follow. Finitists seem to want more: that recipes be guaranteed to output on any admissible input. Given their nominalistic leanings, certain constructivists identify operations with explicit rules or linguistic expressions of rules.

Constructive logics. Constructivism almost always calls for radical revisions in conventional mathematics. E may require, of any statement, that its admissibility entail the availability of constructions. To see why, take a statement S and define a predicate $A(x)$ so that $A(x)$ is true of 0 alone, if S holds, and of 1 alone otherwise. Trivially, when S holds, there is an m such that $A(m)$ and, given E , there must be a construction of such an m . But there is more to see from this example. If the law of the excluded middle (LEM) is valid, there is a simple proof of $\exists n A(n)$. But, if the truth or falsity of S is unknown, this simple proof appears to violate E . Short of settling S , one cannot construct a well-defined m such that $A(m)$, if ‘construction’ and ‘well-definedness’ are interpreted restrictively, since such an m equals 0 just in case S is true. Therefore, E is so interpretable as to conflict with the validity of LEM.

It has also been said that constructivism requires changes in logic for the reason that mathematical claims, constructively treated, hold extra information in the form of recipes for construction. Supplying this information is a requirement on any constructively correct proof. Presumably, the proof of $\exists n A(n)$ just described does not bear appropriate information when S is unknown. In this regard, it is important to note that the impact of E upon logic should not be thought wholly restrictive. Constructive claims can feature in proofs not only as conclusions but also as premises and, hence, as yielding up extra information. Some constructivists, for example, read E^* as underwriting logical principles for deducing, directly from $\forall x \exists y$ premises, axioms of choice.

One way constructivists delimit needed alterations to logic is by establishing canons for correct constructive proof. Given such canons, a principle of inference is to be allowed when it licenses inferences that preserve correct provability. Intuitionists and Russian constructivists, for example, require of disjunctive and universal claims that ($E \vee$) a constructive proof of $(A \vee B)$ include an effective marker and a further constructive proof. If the marker has value 0, the further proof is one of A and, if the marker is 1, the proof is of B and ($E \forall$) a constructive proof of $\forall n A(n)$ involve an operation f such that, for any number m , $f(m)$ yields a constructive proof of $A(m)$.

The basic idea of using canons of abstract proof to assess logical principles is apparent in the writings of Brouwer. A. Heyting and A.N. Kolmogorov were the first to devise both proof specifications for each form of mathematical statement and formal logics arguably correct with respect to them. (Kolmogorov’s logic was not complete, however.) Allegiance to the logic of Heyting, if not to any particular formalization of it, is shared by many constructivists.

Logics obeying $E \vee$ and $E \forall$ may fail to certify LEM. To see this, let $A(n)$ be an undecidable predicate of natural numbers. Then, $\forall n (A(n) \vee \sim A(n))$ plausibly lacks a constructive proof. For, were there such a proof, it would yield a recipe f giving, computably in n , a marker $f(n)$ signalling, in each case, the truth or the falsity of $A(n)$ and contradicting the supposition that $A(n)$ be undecidable.

2 Russian constructivism

Prominent among Russian (or Markovian) constructivists have been A.A. Markov, who initiated the approach around 1950, N.A. Shanin and G.S. Ceitin. As constructivists, the Russians might be thought most exacting, since E is imposed on existential quantifications over any domain. Entities are well-defined only if they admit encoding

as natural numbers or as words on a finite alphabet. Encodings must be, as Markov wrote, ‘potentially realizable’ as concrete notation (1962: 2). Accordingly, Russian constructivists require operations to be given as Turing machines or Markov algorithms and, hence, as finite words. Sets are identified with arithmetic formulas with one free variable.

In virtue of their views on operations, a belief in (a version of) Church’s thesis (CT) is ascribed to Russian constructivists. CT’s most straightforward version is ‘Every total numerical function is Turing computable’. This CT is not, as in recursion theory, a hypothesis about the meaning of ‘informally computable’, but a powerful mathematical axiom. For example, CT entails, in Russian constructivism, the falsity of the general LEM. The Markovian implementation of E, then, enjoins upon Russian constructivists markedly anti-classical principles of logic and mathematics.

Unlike new constructivists and Brouwerian intuitionists, Russian constructivists adopt (a form of) Heyting’s logic as an official formal system. They employ an understanding of the system’s signs devised by Shanin on which the truth of an existential claim S records the existence of an algorithm on whose operation S is a specification. In the name of this understanding, mathematical principles including CT get adopted which are permitted by neither intuitionists nor new constructivists. Among these is Markov’s principle:

(MP) Whenever a natural number predicate $A(x)$ is decidable, $\exists n A(n)$ is provable constructively from $\sim \forall n \sim A(n)$.

When MP is rendered in algorithmic terms, a motive for adopting it is plainest. The decidability of $A(x)$ then requires a recipe R for determining membership in $\{n: A(n)\}$. If one runs R successively on inputs $0, 1, 2, \dots$ searching for members of $\{n: A(n)\}$, a proof of $\sim \forall n \sim A(n)$ guarantees that this search cannot fail. MP says that, in these circumstances, there is an m such that $A(m)$ to which the search procedure will lead.

Adherence to CT lends Russian constructive analysis a computable cast throughout. As O. Aberth wrote, the mathematics ‘may be informally described as an analysis wherein a computation algorithm is required for every entity employed. The functions, the sequences, even the numbers of computable analysis are defined by means of algorithms. In this way definition and evaluation are made inseparable’ (1980: 2). There is no single conception of real number within Russian constructivism. A real is an ‘FR-number’, for example, when it is given in terms of a pair of algorithms, the first yielding a Cauchy sequence of rational approximants and the second a modulus of convergence. (‘FR’ stands for ‘fundamental sequence with regulator of convergence’.) The most famous mathematical theorem of Russian constructivism is that proved by Ceitin (and, independently, by Kreisel, Lacombe and Schoenfield), which implies that every function from the reals to the reals is continuous. In Russian constructivism, this does not entail, as it would conventionally, that every real-valued function on a closed interval is uniformly continuous. Using a result of E. Specker, Russian constructivists prove that there are continuous functions which are unbounded on closed intervals.

3 New constructivism

New constructivists (or Bishop constructivists) comprise a relatively small group of mathematicians inspired by E. Bishop. Bishop’s influential *Foundations of Constructive Analysis* banished lingering doubts that a significant mathematics could survive constructivization. It was consequent to the book’s 1967 publication that new constructivism had its greatest impact upon mathematics.

As for well-defined entities, new constructivists do not require that everything be coded as integers. In addition to numbers, they embrace abstract operations, functions (akin to extensional operations), proofs and sets. There is canonical number notation but no special notations for sets or operations. A set S is said to be ‘constructed’ once recipes are given for constructing S -elements and conditions are specified under which equalities between those elements can be proved. For example, a new constructivist real number is a triple $\langle r, p, s \rangle$ of operations r and s and a proof p wherein r is a sequence of rationals, s a sequence of natural numbers and p a proof that s is a modulus of convergence for r .

As far as constructions are concerned, only mathematical operations representable by demonstrably computable recipes are allowed. Even so, new constructivists refuse CT and set no a priori limit on the way recipes must be expressed. Nor do new constructivists endorse a particular formalism for constructive logic. Certain working

limitations are, however, discernible from a metamathematical viewpoint: CT has been proven consistent with systems reflecting Bishop's ideas. Metamathematical experience also suggests that all new constructive theorems are formally provable from suitable axioms via Heyting's logic.

Bishop insisted that new constructive mathematics capture the 'numerical meaning' of mathematical claims. Numerical meaning, he averred, is the sole clear and legitimate meaning for any mathematics. This meaning was to be cashed out in predicting the outcomes of performable computations on the integers, computations that come from realizing constructive theorems as computer programs. Bishop encouraged a comparison between formalizations of new constructivism and high-level specification and programming languages, ones whose proofs could be compiled into implementable code. This general idea of 'programming constructive mathematics' has since been taken up by a number of computer scientists.

The principal effort in new constructivism, which Bishop called 'the most urgent task of the constructivist' (1970: 54), is not a wholly autonomous construction of mathematics under the guidance of E but a painstaking reconstruction of classical mathematics intended to reveal its numerical meaning. The goal is to set as many traditional theorems as possible onto a constructive basis without offending conventional sensibilities. Conflict never erupts between theorems of new constructivism and those of conventional mathematics since new constructivists shun any anti-classical consequences of E.

Without benefit of nonclassical assumptions, new constructivists cannot prove, as their Russian colleagues do, that every real-valued function of a real variable is continuous. Moreover, Specker's theorem shows that they cannot expect to prove that continuous functions are bounded in range, even on closed unit intervals. In order to retain the boundedness theorem for continuous functions on such intervals, new constructivists supplement the traditional statement of the theorem with the extra hypothesis that the function is uniformly continuous. This kind of hypothesis supplementation has become a feature of all branches of new constructive mathematics.

4 Brouwerian intuitionism

The most celebrated of contemporary mathematical constructivisms, at least before 1967, was Brouwerian (Dutch or neo-) intuitionism. To an extent, the celebrity of intuitionism was allied to that of its leading exponent, L.E.J. Brouwer, who first presented the approach in articles appearing between 1907 and 1930 (see [Intuitionism §§2-3](#)). Brouwer and his followers maintained a relative liberality in the matter of well-definedness. Besides natural numbers, intuitionists granted full status to non-algorithmic or 'free choice' sequences, kinds of sets, abstract proofs (including infinitary ones) and moments of time. Brouwer refused nominalistic constraints: mathematical existence for him did not await the fashioning of notations. Mathematics was to be constructed as a free creation, unaided by realist metaphysics, linguistic representation or empirical data. Brouwer deemed an entity legitimate if it was apparent to a Neo-Kantian inner intuition. A well-defined entity could be displayed in intuition either as a mental construction from the perception of a passage in time or as a choice sequence or set of items so constructed.

From among a panoply of legitimate operations on numbers, intuitionists denominate as 'law-like' those functions which are computable by recipe. In keeping with his anti-linguistic bent, Brouwer did not establish strict limits on the means by which recipes are to be communicated and CT for law-like sequences is not intuitionistically acceptable. Brouwer expressed a concern (one which both Bishop and Markov later proved unfounded) that the law-like continuum - in which reals are identified with law-like sequences - would not support a satisfactory real analysis. This concern seems to have led him to import into intuitionistic mathematics an analogue of the probabilistic 'choice sequences' of Borel. For Brouwer, choice sequences need not be law-like or even fully determinate. There may be some restrictions set upon sequence terms to appear in future but, beyond that, nothing about the course of a sequence may be presumed known.

Traditionally, there are two intuitionistic analogues to the conventional notion of set: 'species' and 'spread'. Species are properties determined by extensional mathematical predicates and individuated extensionally: pairs of species are identical when they have the same members. Spreads, on which Brouwer laid great emphasis, are either species of choice sequences constrained to obey a 'spread law' or species of sequences obtained from those via continuous mappings. The generators of real numbers comprise a spread in the latter sense.

The quintessential intuitionistic result is 'Brouwer's theorem': every real-valued function of a real variable on a closed, bounded interval is uniformly continuous. Brouwer concluded from this that every real-valued function on

the reals is continuous. Brouwer's proof of the latter is markedly different from that which leads to Ceitin's theorem in Russian constructivism; it relies upon mathematical principles inconsistent with CT (see [Intuitionism §3](#)).

5 Finitism

As the name suggests, finitism requires that one eschew reference to infinite collections. In the case of Hilbert's finitism, this requirement was coupled with a desire, epistemically motivated, to replace abstract concepts with concrete, visualizable notation. Hilbert's idea was that a metamathematics confined to the realm of the visualizable would secure conventional mathematics against the paradoxes. Besides Hilbert's finitistic metamathematics, notable forms of twentieth-century finitism are represented by Skolem's primitive recursive arithmetic and Yessenin-Volpin's ultra-intuitionism.

For most finitists, only natural numbers or items encodable as natural numbers count as fully well-defined. Numbers are not conceived as constituting an infinite totality but as individually realizable in concrete notation. Hilbert would have had it that talk of natural number, finitistically construed, comes down to talk of (the intuitable forms of) strings of tally marks. He wrote, 'The subject matter of mathematics is, in accordance with this theory, the concrete symbols themselves whose structure is immediately clear and recognizable' ([1926] 1983: 142) (see [Hilbert's programme and formalism](#)).

If the use of unbounded existential quantification is believed to require infinite totalities, as it was by Hilbert, restriction E will be implemented finitistically using constructions carrying explicit bounds. Bounded existential quantifications over the natural numbers are usually finitistically admissible and other existentially quantified claims, when allowed, get treated as 'incomplete statements', abbreviations for bounded quantifications. Universal quantifications are, if admissible, finitistically reconstrued as free variable expressions each instance of which is verifiable or falsifiable by direct calculation.

Constructions do enter into finitistic mathematics, but not as members of an infinite domain of number operations, as in intuitionism. Nor are they identified with members of a category of abstract rules or infinite graphs. An operation on natural numbers is a finitistic construction if realizable as a combinatorial transformation of concrete signs for which explicit bounds on the number of steps required to complete the transformation have been given. Primitive recursive functions are natural candidates for finitistic constructions on the natural numbers since these functions are Turing computable functions definable without unbounded search routines. Whether the class of finitistically admissible functions includes more than the primitive recursive functions is a topic of disagreement.

In Hilbertian finitism, admissible statements are said to capture mathematical claims which are '*inhaltlich*' or contentual. Statements refusing finitistic reconstrual were considered not contentual but ideal, expressions lacking full content. For Skolem, only those statements whose truth-values are fully determinable via primitive recursive calculations are admitted. Arithmetical predicates which are composed of simple equations between primitive recursive terms combined by connectives and bounded quantifiers Skolem treated as primitive recursive functions. Therefore, systems representing this approach are called 'logic-free', since quantifiers and connective signs are not required. Because of the limitations on the kinds of claim permitted, the proofs in Skolem's and Hilbert's finitistic arithmetics obeyed conventional logic.

Finitistic number theory and analysis, in the style of Skolem, has undergone some development, notably by R. Goodstein. Various constructions of real number are examined, among them that of primitive recursive real number: a primitive recursive sequence armed with a primitive recursive modulus. Anti-classical mathematical axioms are refused; hence, there are no finitistic continuity theorems like Ceitin's. However, since finitistic mathematics is intuitionistically acceptable, neither will it be possible to contradict Brouwer's theorems. According to Goodstein, a finitist in the style of Skolem cannot prove that a bounded, monotone sequence of rational numbers is convergent.

The so-called 'ultra-intuitionism' of A.S. Yessenin-Volpin differs radically from the views just described. Standard concepts of natural number and Skolem's 'recursive mode of thought' are not accepted as unequivocally correct. Yessenin-Volpin limits well-definedness to numbers which are feasible rather than potentially realizable: to be well-defined, an ultra-intuitionistic number must be literally displayed as a notation. One might say that the well-defined numbers of ultra-intuitionism do not constitute a single potentially infinite series but form a tree, each

branch of which represents a temporally expanding series of tokens from some numeral system. From P. Bernays and G. Kreisel, commentators have adopted the term ‘strict finitism’ for application to such positions as Yessenin-Volpin’s in distinguishing them from more liberal finitistic conceptions. Among philosophers of language, the viability of strict finitism has proved a subject of controversy. M. Dummett (1975) has maintained that sorites-like paradoxes would so haunt strict finitism that it is revealed as incoherent; C. Wright (1982) has since raised serious questions about the correctness of Dummett’s conclusions.

6 Predicativism

Predicativism took impetus from the writings of Poincaré and Russell on the paradoxes and received influential, but troubled, expression in Whitehead and Russell’s *Principia Mathematica* (1910-13) (see [Theory of types](#)). Predicativistic approaches to the foundations of mathematics were later instigated or investigated by C. Chihara, S. Feferman, G. Kreisel, P. Lorenzen, H. Wang and H. Weyl, among others.

Predicativists would have it that all mathematical entities - or all higher-order entities - are logical classes and that their existence is governed by a version of E. A logical class $\{x: A(x)\}$ is predicatively well-defined on condition that all quantifiers in $A(x)$ are restricted to a range of classes or entities already known to be well-defined and including neither $\{x: A(x)\}$ nor any class presupposing it for its definition. Definitions $\{x: A(x)\}$ satisfying this condition are termed ‘predicative’ and others ‘impredicative’. Russell put forward the requirement that all classes be predicative in the form of what came to be known as the ‘vicious-circle principle’. According to Russell, predicative classes must be organized into types so that ‘any expression containing an apparent [bound] variable is of higher type than that variable’ ([1908] 1967: 182). Consequently, the favoured way to implement bans against impredicativity has been to devise hierarchical formal systems (predicative type theories) for class definition and manipulation.

An intended model of predicative formalism, greatly simplified, is segmented into an ascending sequence of orders or levels indexed by natural or ordinal numbers. At the lowest or 0 level is a collection of given elements such as the natural numbers. Level 1 contains those classes whose definitions involve quantifiers ranging only over the given elements, level 2 contains classes specifiable using quantifiers ranging either over the given elements or over classes of level 1, and so on. Each class has a level, so impredicative classes are ruled out. Formalisms embodying an idea of levels include the ‘ramified type theory’ of *Principia Mathematica* and Wang’s systems Σ .

A class is predicative thanks to a suitable definition or notation for it; that notation cannot be grasped before notations for classes of lower level are grasped. As Wang put it, ‘new objects are only to be introduced stage by stage without disturbing the arrangement of things already introduced or depending for determinedness on objects yet to be introduced at a later stage’ (1970: 640). Following Poincaré, one often imagines predicative universes as expanding over time.

As for constructions, we can think of each predicative class as the result of a well-founded construction process involving conventional logical or class operations such as complementation, intersection and union, and quantification which is predicatively restricted. Hence, the imposition of E upon predicative mathematics seems to call for no great alteration in the fundamental inferences of conventional higher-order logic, apart perhaps from variable indexing. Each variable may wear an ordinal index and quantifier rules be so restricted that bound variables share indices with substituted parameters.

Predicativism has not spawned a programme for reworking mathematics as influential as Brouwerian intuitionism or new constructivism. Even so, a wide variety of mathematical results are known. Weyl saw that the theorem that every non-empty bounded class of reals has a least upper bound is unobtainable in complete generality within (a restricted) predicative mathematics if reals get defined in terms of classes. He showed that, by treating reals as sequences, the theorem that every bounded, monotonically increasing sequence approaches a limit is predicatively provable, as are other theorems of analysis. Wang has argued that predicativists cannot retain Cantor’s theorem on the existence of uncountable sets.

See also: [Logical and mathematical terms, glossary of](#)

DAVID CHARLES McCARTY

References and further reading

- Aberth, O.** (1980) *Computable Analysis*, New York: McGraw-Hill. (An attractive technical introduction to analysis in the style of Russian constructivism.)
- Beeson, M.J.** (1985) *Foundations of Constructive Mathematics*, Berlin: Springer. (An overview of the metamathematics of constructive formal systems. Requires considerable formal experience.)
- Benacerraf, P. and Putnam, H.** (eds) (1964) *Philosophy of Mathematics: Selected Readings*, Englewood Cliffs, NJ: Prentice Hall, 2nd edn, 1983. (An invaluable collection of classic essays on the foundations of mathematics; most are nontechnical.)
- Bishop, E.** (1967) *Foundations of Constructive Analysis*, New York: McGraw-Hill. (The bible of new constructivism. A technical work with a lucid nontechnical introduction.)
- Bishop, E.** (1970) 'Mathematics as a Numerical Language', in J. Myhill, R.E. Vesley and A. Kino (eds) *Intuitionism and Proof Theory*, Amsterdam: North Holland, 53-71. (A clear statement of Bishop's central philosophical ideas.)
- Bridges, D. and Richman, F.** (1987) *Varieties of Constructive Mathematics*, Cambridge: Cambridge University Press. (A mathematical introduction to constructivism, contrasting theorems of new constructivism with those of intuitionism and Markovian constructivism. A background in conventional analysis is required.)
- Brouwer, L.E.J.** (1981) *Brouwer's Cambridge Lectures on Intuitionism*, ed. D. van Dalen, Cambridge: Cambridge University Press. (A compact sample of Brouwer's thinking; well edited but quite technical.)
- Chihara, C.** (1973) *Ontology and the Vicious-Circle Principle*, Ithaca, NY: Cornell University Press. (Includes an exposition, for philosophers, of traditional issues surrounding predicativity.)
- Dummett, M.A.E.** (1975) 'Wang's Paradox', *Synthese* 30: 301-24; repr. in *Truth and Other Enigmas*, Cambridge, MA: Harvard University Press, 1978, 248-68. (A thought-provoking exploration of issues from philosophy of language which may affect the legitimacy of finitism.)
- Dummett, M.A.E.** (1977) *Elements of Intuitionism*, Oxford: Clarendon Press, 1990. (An influential text in which great emphasis is placed upon the metalogic of intuitionism.)
- Feferman, S.** (1964) 'Systems of Predicative Analysis', *Journal of Symbolic Logic* 29: 1-30. (A brilliant introduction to predicativism and its systematic formulation. The second section requires technical knowledge.)
- Hazen, A.** (1983) 'Predicative Logics', in D. Gabbay and F. Guenther (eds) *Handbook of Philosophical Logic*, vol. 1, *Elements of Classical Logic*, Dordrecht: Reidel. (A helpful overview of and introduction to predicativism and predicative logics.)
- Heijenoort, J. van** (ed.) (1967) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879- 1931*, Cambridge, MA: Harvard University Press. (An invaluable collection of significant articles by Gödel, Hilbert, Brouwer, Skolem and others. Technical knowledge often required.)
- Heyting, A.** (1956) *Intuitionism: An Introduction*, Amsterdam: North Holland; 3rd revised edn, 1971. (The first - and still the most charming - popular technical account of Brouwer's intuitionism.)
- Hilbert, D.** (1926) 'Über das Unendliche', *Mathematische Annalen* 95: 161-90; trans. 'On the Infinite', in P. Benacerraf and H. Putnam (eds) *Philosophy of Mathematics: Selected Readings*, Cambridge: Cambridge University Press, 2nd edn, 1983; and in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic 1879-1931*, Cambridge, MA: Harvard University Press, 1967. (A classic statement of Hilbert's views on metamathematics; largely nontechnical.)
- Markov, A.A.** (1962) 'O Konstruktivnoii Matematike', *Trudy Matematicheskogo Instituta imeni V.A. Steklova* 67: 8-14; trans. 'On Constructive Mathematics', *American Mathematical Society Translations*, series 2, 98: 1- 9, 1971. (A brief introduction to Russian constructivism; relatively nontechnical.)
- Russell, B.A.W.** (1908) 'Mathematical Logic as Based on the Theory of Types', *American Journal of Mathematics* 30: 222-62; repr. in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic 1879-1931*, Cambridge, MA: Harvard University Press, 1967, 150-82. (Includes Russell's diagnosis of the paradoxes and an explanation of his ramified type theory.)
- Skolem, T.** (1923) 'Begründung der elementaren Arithmetik durch die rekurrierende Denkweise ohne Anwendung scheinbarer Veränderlichen mit unendlichem Ausdehnungsbereich', *Skifter utgit av Videnskapsselskapet I Kristiania, I. Matematisk-naturvidenskabelig Klasse* 6: 1-38; trans. 'The Foundations of Elementary Arithmetic Established by Means of the Recursive Mode of Thought, Without the Use of Apparent Variables Ranging Over Infinite Domains', in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic 1879-1931*, Cambridge, MA: Harvard University Press, 1967, 302-33. (A brilliant and careful formal development of a finitistic arithmetic.)
- Tait, W.W.** (1981) 'Finitism', *Journal of Symbolic Logic* 78 (9): 524-46. (A lucid exposition of one approach to the foundational issues surrounding finitism.)
- Troelstra, A.S. and Dalen, D. van** (1988) *Constructivism in Mathematics: An Introduction*, Amsterdam: North

Holland, 2 vols.(A comprehensive technical overview of the current state of constructive mathematics, with emphasis on intuitionism and mathematical logic.)

Wang, H. (1970) *Logic, Computers and Sets*, New York: Chelsea.(Includes several articles by Wang on the historical and technical background to predicativism.)

Weyl, H. (1918) *Das Kontinuum: Kritische Untersuchungen über die Grundlagen der Analysis*, Leipzig: Veit; trans. T. Bole, *The Continuum: A Critical Examination of the Foundations of Analysis*, New York: Dover, 1994.(Historically influential but, at times, idiosyncratic treatment of mathematical and logical foundations for predicative analysis.)

Whitehead, A.N. and Russell, B.A.W. (1910) *Principia Mathematica*, vol. 1, Cambridge: Cambridge University Press; 2nd edn, 1925.(The classic effort to demonstrate the logicistic thesis; highly technical.)

Wright, C. (1982) 'Strict Finitism', *Synthese* 51; repr. in *Realism, Meaning and Truth*, Oxford: Blackwell, 1986, 107-75.(A detailed reconsideration of issues raised in Dummett (1975).)

Continuum hypothesis

The ‘*continuum hypothesis*’ (CH) asserts that there is no set intermediate in cardinality (‘size’) between the set of real numbers (the ‘*continuum*’) and the set of natural numbers. Since the continuum can be shown to have the same cardinality as the power set (that is, the set of subsets) of the natural numbers, CH is a special case of the ‘*generalized continuum hypothesis*’ (GCH), which says that for any infinite set, there is no set intermediate in cardinality between it and its power set.

Cantor first proposed CH believing it to be true, but, despite persistent efforts, failed to prove it. König proved that the cardinality of the continuum cannot be the sum of denumerably many smaller cardinals, and it has been shown that this is the only restriction the accepted axioms of set theory place on its cardinality. Gödel showed that CH was consistent with these axioms and Cohen that its negation was. Together these results prove the independence of CH from the accepted axioms.

Cantor proposed CH in the context of seeking to answer the question ‘What is the identifying nature of continuity?’. These independence results show that, whatever else has been gained from the introduction of transfinite set theory - including greater insight into the import of CH - it has not provided a basis for finally answering this question. This remains the case even when the axioms are supplemented in various plausible ways.

1 The continuous, the discrete and the infinite

Greek mathematics was divided into arithmetic - the science of discrete magnitudes - and geometry - the science of continuous magnitudes. The basis for this division can be found in Aristotle’s sophisticated discussions of space and time. The existence of incommensurable magnitudes, such as the side of a square and its diagonal (the ratio of which is $\sqrt{2}$), was already well known, as was the distinction between points and extended parts, however small, of a continuous whole. Aristotle argued that since two points can never be contiguous (either there is a separation between them or they coincide), a continuum can never be built up out of points. He defined a continuously extended whole as one which may be divided at any point into continuously extended wholes and concluded that it is necessary to draw a fundamental distinction between those wholes which are collections of parts and those which are ‘given before their parts’, that is, between discrete and continuous magnitudes.

So long as this division was in place (that is, until the seventeenth century) a question such as ‘What is the (cardinal) number of points in a continuous line?’ could not arise because it could not sensibly be asked. ‘Continuity’ was a primitive geometric concept. The situation changed, however, with the acceptance of Descartes’ introduction of algebraic methods into the heart of geometry, allowing geometric curves to be defined by equations of ‘motions’ of points ‘generating’ them. This effectively swept aside the division between arithmetic and geometry, but without in any way confronting or resolving the problem of how to model the continuous by the discrete.

Newton and Leibniz explored and exploited the infinite divisibility of continua in their attempts to study continuous motions algebraically. Initially this exploration (infinitesimal calculus, analysis; see [Analysis, philosophical issues in §1](#)) moved freely between algebraic expression and geometric intuition, mediated by the concept of continuous motion. However, this led to the discovery of a number of ‘pathological’ functions, whose graphs cannot be pictured. For example, in a continuous plane, unlimited in extent, there is no limit to the number of wavelike, and hence continuous, oscillations that a point might execute within a given finite interval. So one might ask whether there might be a function which corresponds to a point oscillating with infinite frequency (that is, which is such that no matter how small an interval we take there will be an oscillation contained within it). Weierstrass’ everywhere continuous but nowhere differentiable function seemed to ‘describe’ such a ‘motion’. This kind of function contradicts the assumption, based on cases which can be pictured, that continuity and differentiability go together. So an account of continuity was required in terms not based on geometric intuition.

2 Numbering the continuum

Cantor’s early attempts to characterize the structure of continua were natural extensions of the kind of work in analysis which led to discovery of these pathological functions. He established that the cardinality of the set $\mathbb{R}$ of real numbers is greater than that of the set $\mathbb{N}$ of natural numbers (see [Cantor’s theorem](#)). He further proved that

there is a one-one correspondence between the points in a plane (the two-dimensional continuum) and the points in a line. So he already knew (writing $|A|$ for the cardinality of A) that

$$|\mathbb{R} \times \mathbb{R}| = |\mathbb{R}| > |\mathbb{N}|.$$

Cantor thus interpreted the problem of how to identify the nature of continuity as that of how to say when a given set of points in an n -dimensional space is to be considered a continuum. Having already defined the real numbers in terms of limits of convergent sequences of rational numbers, he made use of the more general notion of a ‘limit point’. A point p is defined to be a limit point of a set P of points if in any neighbourhood of p there are infinitely many points of P . It can be proved that if P is infinite then it has at least one limit point. The set of limit points of P is denoted by $P^{(1)}$. The set $\mathbb{R}$ of real numbers (the ‘continuum’) is such that $\mathbb{R}^{(1)} = \mathbb{R}$. Sets with this property were called ‘perfect’. But perfection does not serve to characterize continuity, as Cantor demonstrated with his ‘ternary set’, also called ‘Cantor’s discontinuum’ (see [Set Theory §5](#)). This is the set of points of the closed interval (a ‘closed’ interval includes the points at either end) $[0, 1]$ which remain after infinitely many repetitions of the following procedure: divide each interval of the set into thirds, and in each case discard (‘blank out’) the middle third, leaving behind two closed intervals.



This set is dense in no interval of the continuum ($B \subseteq A$ is ‘dense’ in A if it is disjoint from no open interval of A), closed, perfect and contains no interior points: it is completely discontinuous.

Nevertheless, Cantor felt that perfect sets held the key since he was able to prove that no *denumerable* point set (that is, a set with the cardinality of the natural numbers) could be perfect. If a set P is finite, its first derived set $P^{(1)}$ is empty. For any P , either there is some finite n such that $P^{(n)}$ is finite and hence $P^{(n+1)}$ is empty, or there is not. For perfect sets there is no such finite n . So if this idea is to differentiate between continuous and discontinuous perfect sets it seems that *transfinite* indices of iteration need to be used. This was the context in which Cantor originally introduced transfinite ordinal numbers (see [Set theory §2](#)). He divided the ordinals into first and second number classes:

- (1) $0, 1, 2, 3, \dots$
- (2) $\omega, \omega + 1, \omega + 2, \dots, \omega \cdot 2, \omega \cdot 2 + 1, \dots, \omega \cdot 3, \dots, \omega \cdot \omega, \dots$

(1) is the set of natural numbers, $\mathbb{N}$, which has cardinality $\aleph_0$ (‘aleph-nought’ or ‘aleph-null’). Every ordinal α in (2) is such that the cardinality of $\{x: x < \alpha\}$ is $\aleph_0$. That is, these numbers represent all the possible orderings of a denumerably infinite set. Cantor proved ([\[1895, 1897\] 1955: 169-73](#)) that these two number classes cannot be put into one-one correspondence and that there can be no set with a cardinality in between the two. This justified calling the cardinality of (2) ‘ $\aleph_1$ ’.

In this way the extension of the ordinal number sequence into the transfinite generates not only infinite ordinal numbers α but also infinite cardinal numbers $\aleph_\alpha$. It thus provides a scale of cardinal numbers on which one might hope to locate $2^{\aleph_0}$. The ‘continuum hypothesis’ (CH) expresses the claim that there is a one-one correspondence between the real numbers and the numbers in the second number class (in symbols, $2^{\aleph_0} = \aleph_1$), and another way of interpreting the claim made by the generalized continuum hypothesis (GCH) would be as saying that the two routes to generating higher infinite cardinalities (via power sets and via the number classes) coincide (in symbols, $2^{\aleph_\alpha} = \aleph_{\alpha+1}$ for every ordinal α). Although Cantor believed CH to be true (and Gödel ([1939](#)) proved it to be consistent with the axioms of set theory), he realized that there is no immediate guarantee that every cardinality will have a representative among the cardinalities of the number classes unless it can also be assumed that every set can be well-ordered, and hence has an ordinal number. If it were not the case that every set could be well-ordered, the cardinalities of those sets which cannot be so ordered would not be represented by cardinal numbers arising from the construction of the ordinal number sequence. Indeed it can be shown that in

Zermelo-Fraenkel set theory (ZF), GCH entails the axiom of choice and hence that every set can be well-ordered, whereas Cohen's independence result proves that the converse does not hold (see Cohen 1963-4).

3 Reactions to independence

Proof of the independence of CH (GCH) from ZFC (ZF plus the axiom of choice - see [Set theory, different systems of §3](#)) stimulated renewed interest in philosophical questions about the status of set theory and the epistemology of mathematics. Is there a universe of sets in virtue of which CH is either true or false, although at present we do not know which? If this is the case, how would we come to know more about it and hence hope to determine the truth-value of CH? The most radical reaction would undercut even these questions by interpreting the independence results as showing that the whole set-theoretic approach to problems of continuity was misguided: CH is undecidable because it does not make any real mathematical sense; it presupposes that we can sensibly talk about actually infinite sets, whereas we cannot. Less radical reactions assume that CH does make mathematical sense. The problem then is not merely how to determine whether or not it should be accepted, but to decide whether there is a basis for acceptance or rejection and, if so, what this might be. It might be argued that the situation of CH is analogous to that of the parallel postulate after it was proved to be independent of the remaining axioms of Euclidean geometry. The result in that case was a widening of the field of mathematical investigation to consider a whole variety of geometries in which the parallel axiom fails. Analogously, in set theory, explorations of systems in which CH holds and in which it fails have been carried out. Gödel (1947), however, contrasts the epistemological situations in geometry and set theory. In geometry the question of the truth or falsity of the parallel postulate retains its sense if the geometric axioms are interpreted as referring to the behaviour of rigid bodies, light rays and so on, but not otherwise - that is, the question of its truth or falsity has become an empirical question to be answered by reference to the contexts in which geometry is used; it is no longer a mathematical question. The question of the truth or falsity of CH, however, retains its sense only if the axioms of set theory are interpreted as referring to mathematical objects (sets), since these axioms have no direct empirical interpretation. Gödel himself believed that sets are mathematical objects which have an independent existence. He would thus interpret the proof of the independence of CH from ZFC as a demonstration that the axioms of ZFC do not contain a complete description of the set-theoretic universe. In this case there is a need to look for new axioms to complete the description. Several such axioms have been suggested and their consequences studied (see [Set theory §1](#)), but none has yielded anything approaching consensus on either their acceptability or that of CH.

Gödel's mathematical realism is not the only philosophical position which would underwrite attempts to resolve questions about the status of CH (and other proposed axioms of set theory) by studying the consequences for other areas of mathematics of accepting or denying it. It might be argued that what mathematical sense CH has is derived from its preformal connections with the problem of characterizing the continuous in terms of the discrete. If so, any determination should be based on its impact on the ever-widening area of mathematics connected with this problem. This approach relates philosophical issues about the status of set theory with other areas of mathematics, it does not resolve them.

See also: [Logical and mathematical terms, glossary of](#)

MARY TILES

References and further reading

- Cantor, G.** (1895, 1897) 'Beiträge zur Begründung der transfiniten Mengenlehre', parts 1 and 2, *Mathematische Annalen* 46: 481-512, 49: 207-46; trans. P.E.B. Jourdain (1915), *Contributions to the Founding of the Theory of Transfinite Numbers*, New York: Dover, 1955. (Two of Cantor's most important memoirs on transfinite numbers.)
- Cohen, P.J.** (1963-4) 'The Independence of the Continuum Hypothesis', *Proceedings of the National Academy of Science, USA* 50: 1,143-8, 51: 105-10; repr. in *Set Theory and the Continuum Hypothesis*, New York: Benjamin, 1966. (The original version of his proof.)
- Dauben, J.W.** (1979) *Georg Cantor: His Mathematics and Philosophy of the Infinite*, Cambridge, MA: Harvard University Press. (An excellent intellectual biography which traces the development of Cantor's mathematical ideas, while putting them in the context of his philosophy of the infinite and his professional life.)
- Gödel, K.** (1939) 'Consistency Proof for the Generalized Continuum Hypothesis', *Proceedings of the National Academy of Science, USA* 25: 220-4. (Original publication of the proof.)

- Gödel, K.** (1940) *The Consistency of the Continuum Hypothesis*, Princeton, NJ: Princeton University Press, 1994. (Revised and more complete version of the original proof.)
- Gödel, K.** (1947) 'What is Cantor's Continuum Problem?', *American Mathematical Monthly* 54: 515-25; repr. in P. Benacerraf and H. Putnam (eds) *Philosophy of Mathematics: Selected Readings*, Cambridge: Cambridge University Press, 2nd edn, 1983. (A very valuable discussion which gives some of the context of debate about CH.)
- Hallett, M.** (1984) *Cantorian Set Theory and Limitation of Size*, Oxford: Clarendon Press. (Includes an integrated account of Cantor's metaphysical and mathematical theories of infinity together with discussion of the relation between Cantorian and modern set theory.)
- Tiles, M.E.** (1989) *The Philosophy of Set Theory: An Introduction to Cantor's Paradise*, Oxford: Blackwell. (Traces the philosophical and mathematical background to the formulation of CH and considers whether it can be decided.)

Conventionalism

How is it known that every number has a successor, that straight lines can intersect each other no more than once, that causes precede their events, and that the electron either went through the slit or it did not? In cases like these it is not easy to find observable evidence, and it is implausible to postulate special modes of intuitive access to the phenomena in question. Yet such theses are relied on in scientific discourse and can hardly be dismissed as meaningless metaphysical excess. In response to this problem the positivists and empiricists (notably Poincaré, Hilbert, Carnap, Reichenbach and Ayer) developed a strategy known as conventionalism. The idea was that certain statements, including fundamental principles of logic, arithmetic and geometry, are asserted as a matter of conventional stipulation, being no more than definitions of some of their constituent terms; consequently they must be true, our commitment to them cannot but be justified, and the facts in virtue of which they are true are simply the facts of our having made those particular decisions about the use of words. This doctrine was a compelling and powerful weapon in the positivist-empiricist arsenal, evolving throughout the 1920s, 1930s and 1940s. But it fell into disfavour under a barrage of serious challenges due mainly to Quine. How are 'conventions' to be identified as such? How could they possibly provide words with meanings, or have the epistemological import that is claimed for them? How could arbitrary, contingent decisions about the use of words result in the existence of necessary facts? In the absence of satisfactory replies to these objections few philosophers these days believe that conventionalism can settle the semantic, epistemological and metaphysical questions that it was intended to answer. However, certain aspects of the view remain defensible and interesting.

1 The scope and nature of conventionalism

Conventionalism is the strategy of supposing that certain theses are implicit definitions - hence mere choices of language - and hoping thereby to explain how their adoption can be justified and to specify that in which their truth consists. It has been applied to the principles of deductive inference, to the claims of mathematics, and to 'bridge laws' specifying the observational consequences of theoretical hypotheses - all domains in which methods of empirical verification might seem to be inappropriate. For example, it was said that the axioms and rules of classical deductive logic are nothing more than implicit definitions of the logical constants, 'and', 'not', 'every', and so on; and this was supposed to show both how our reasoning could be justified and to articulate what it is to be a logical truth. Another famous application was to questions about the structure of space. It was maintained by H. Poincaré (1905) that geometrical facts are matters of convention: thus if we decide to use Euclidean geometry, we thereby implicitly define the terms 'point' and 'line' in such a way that the theorems of Euclidean geometry are true. We are therefore justified in employing any geometry we please; for there are no facts about the structure of space that could make some choices correct and others incorrect. A variant form of conventionalism about space and time was urged by Reichenbach (1968). In his view what stand in need of stipulation are certain correlations between spatiotemporal and physical properties (for example, that light travels in straight lines, and that pendulum swings consume equal periods of time). The idea here was that only once such 'coordinative definitions' had been decided upon would it become possible to investigate empirically the structure of space and time; alternate conventions will yield alternate answers, and so there are no facts of the matter as to which answers are correct (see Reichenbach, H. §4; *Relativity theory, philosophical significance of §3*).

Thus, to whatever domain it was applied, conventionalism in its standard form consisted in the following four elements: one, that some specified, determinate body of sentences (or rules) is accepted as a matter of conventional stipulation; two, that these commitments implicitly define certain constituent terms; three, that the propositions which the sentences express are known a priori; and four that those propositions are true - and the corresponding facts exist - in virtue of our conventional stipulations. In the following sections we shall consider objections to each of these components of the doctrine.

2 The definition of 'convention'

In the ordinary sense of the word a *convention* is an explicit agreement to act in some specified way. But, in that sense, no convention has ever been adopted to go around saying ' $1 + 1 = 2$ '; and similarly for the other cases in which conventionalism was deployed. Thus emerges one of the fundamental difficulties faced by this philosophy: that of explaining how to demarcate 'conventions' or 'stipulations' (in the intended technical sense) from other

utterances we are inclined to make. R. Carnap (1950), perhaps the most sophisticated practitioner of the strategy, equated 'linguistic conventions' with the analytic (framework) sentences that determine what our words mean and which language we are using. These were in turn identified as the sentences which we decide to accept on the basis of pragmatic rather than epistemological considerations, and it was concluded that the propositions expressed by those sentences are known a priori. However, Quine's (1935, 1951, 1954) influential critique of conventionalism made a strong case that although these conceptions of 'convention', 'meaning', 'language', 'pragmatic reason', 'analytic' and 'a priori' may well be linked in the way that Carnap supposed, they are all equally obscure and equally in need of clarification, and that it remains to find a way of breaking out of the circle of inter-definable ideas in order to characterize any one of them objectively (see Quine, W.V. §8).

3 The semantic component of conventionalism

The legitimacy of *deciding* that a certain set of sentences (or rules of inference) is to qualify as correct rested on two assumptions: first, that we are at liberty to decide which language to speak, for example, what our words are to mean; and, second, that in stipulating that a sentence be true we are merely fixing the meaning of one or more of its constituent terms. But the second of these assumptions has received considerable criticism, notably in the writings of Frege (1903, 1906), Prior (1960), Belnap (1962) and Dummett (1977). The thesis in question tends to be construed as follows: that in deciding to regard 'Čnum;F' as true, one is implicitly deciding to give 'F' whatever meaning it would need to have in order that 'Čnum;F' be true, given the already established meaning of the rest of the sentence, 'Čnum;_'. But this thesis is far from obviously correct. In the first place, how can we be sure that there exists *any* meaning which could be assigned to 'F' and which would render 'Čnum;F' true? Second, how can we be sure that there is only *one* such meaning? Third, even if there is such a meaning, what reason is there to suppose that this is the meaning that 'F' acquires? And fourth, even if it is, how does our treating 'Čnum;F' as true *bring it about* that 'F' acquires that meaning: what account of the nature of meaning could explain how this happens? The difficulty of answering these questions casts doubt on the possibility of implicit definition.

4 The epistemological component of conventionalism

The main point of the doctrine was to show how our views about logic, arithmetic, spatiotemporal structure, and theory-observation relations could be rational, despite the absence of intuitive or empirical access to those domains. The idea was that our claims are justified a priori - independently of intuition or experience - because they merely fix the meanings of terms, which are commitments that are not subject to a posteriori constraint. Now we have just seen how the feasibility of implicit definition may well be questioned. But there is an additional problem concerning the alleged non-empirical character of such conventions. For even if it is conceded that by regarding a sentence as true one can fix the meaning of one of its constituents, there can remain a doubt as to whether the truth of that sentence is known a priori. The grounds for this doubt derive from Quine's web-of-belief model of theoretical development, according to which our total theory formulation evolves under the constraints of empirical adequacy, global simplicity and conservatism (see Quine, W.V. §3). For if this model is correct then every element in the web earns its keep empirically by contributing towards the overall system that 'best' satisfies these three desiderata, and so everything we believe is a posteriori. This is not to deny that our commitment to 'Čnum;F' might implicitly define a constituent term 'F'. The point rather is that given Quine's model not even such meaning-constituting principles would be known a priori (see A priori).

5 The metaphysical component of conventionalism

On the metaphysical front, the standard conventionalist stance is blatantly antirealist; for insofar as a certain fact is merely a matter of convention, it evidently does not exist independently of us and our practices (see **Scientific realism and antirealism**). And this metaphysical thesis has provided a glaring target for criticism. For although it is no doubt a fact about us that we decide to commit ourselves to the truth of, for example, '1 + 1 = 2', this fact regarding our commitment is not plausibly identifiable with the arithmetical fact that 1 + 1 = 2. The former fact concerns words, the latter numbers. Moreover, the former is contingent - there are possible worlds in which we decide to adopt different conventions - whereas the latter is necessarily the case - there is no possibility that 1 + 1 = 3. Thus the conventionalist's metaphysical thesis, implying as it does that if we had spoken a different language then perhaps 1 + 1 would have been equal to 3, is extremely unattractive and has helped bring the entire

doctrine into disrepute.

6 Prospects for the defence of conventionalism

The above criticisms need not be considered decisive. In each case replies might be devised, involving modifications of the standard four-pronged doctrine. Thus, in response to the challenge to demarcate the class of conventional stipulations, one might abandon the intuitive concept of 'convention', with the presumption that it is evident when such things occur, and look to empirical linguistics for a characterization of which regularities in the use of a term (including assertions containing it) are explanatorily basic and hence implicitly define it. As for the difficulties with implicit definition, one might hope to pre-empt them by altering the assumption that, in regarding 'Čnum;F' as true, one is giving 'F' the meaning it would need to have for 'Čnum;F' to be true. Instead, again invoking a use-conception of meaning, one might suppose that, in regarding 'Čnum;F' as true, one is giving 'F' the meaning that is constituted by that rule of use (Horwich 1997). In that case the four above-mentioned difficulties could not arise. For the existence and uniqueness of a meaning is ensured by the existence of the regularity, and the explanation of how the convention provides meaning would be trivial.

Turning to the epistemological objection, one might resist the thesis that *everything* is a priori, which seemed to follow from the web-of-belief model. For there can surely be different formulations of the same theory - notational variants of each other. Therefore the decision to adopt one rather than another way of expressing one's beliefs is a convention that is not subject to a posteriori revision. Thus the acceptance of a theory, Tabc, formulated using new terms 'a', 'b' and 'c', may be regarded as the product of two commitments: one being the substantive existential belief that there are unique theoretical properties, x, y and z, satisfying Txyz; and the other being the linguistic decision to call those properties 'a', 'b' and 'c'. This conditional thesis, $(\exists!xyz)(Txyz) \rightarrow Tabc$, would not be held a priori (Lewis 1970; Horwich 1986).

Finally, the metaphysical critique of conventionalism could be accepted and accommodated by jettisoning the antirealist component of the standard doctrine. It would still be supposed that certain sentences ('conventions') implicitly define certain of their constituents; but it would no longer be maintained that the facts expressed by those sentences are reducible to facts about our practices, hence matters of convention.

Thus one might respond to the variety of objections to conventionalism. But this can hardly qualify as a vindication of the doctrine. For the revised defensible position which the objections have compelled solves none of the epistemological or metaphysical problems for which conventionalism was invented. Moreover, there is now reason to suspect that there were no real problems here in the first place. For Quine's model suggests that logic, geometry, arithmetic and our theory-observation bridge principles have an epistemological and metaphysical status that is no different from the rest of what we believe. This still leaves open the possibility of implicit definition: the conventional a priori decision to use these words rather than those in the articulation of our beliefs, including those of logic, mathematics and science. But this is such a pale shadow of the original positivistic idea, it might be as well to reserve the term 'conventionalism' for that doctrine and consider it refuted.

See also: [Language, conventionality of](#); [Logical positivism](#); [Necessary truth and convention](#)

PAUL HORWICH

References and further reading

- Ayer, A.J.** (1936) *Language, Truth and Logic*, New York: Dover. (Forcefully advocates conventionalism and elaborates its extensive philosophical ramifications.)
- Belnap, N.D.** (1962) 'Tonk, Plonk and Plink', *Analysis* 22, 130-3. (Discusses implicit definition and the question of whether a *unique* meaning is determined. See §3.)
- Carnap, R.** (1950) 'Empiricism, Semantics and Ontology', in *Meaning and Necessity*, Chicago, IL: University of Chicago Press; 2nd edn, 1956. (The most detailed and plausible articulation of pre-Quinean logical empiricism. See introduction and §2.)
- Dummett, M.** (1977) *Elements of Intuitionism*, Oxford: Clarendon Press. (Argues that the logical constants are not implicitly defined by the laws in which they appear, and hence that classical logic cannot be justified by regarding its fundamental principles as implicit definitions. See §3.)
- Frege, G.** (1895-1903) 'Correspondence between Frege and Hilbert', in *Philosophical and Mathematical Correspondence*, Chicago, IL: University of Chicago Press, 1986, 31-52. (Contains Hilbert's idea - and

Frege's criticism of it - that the axioms of a theory can fix the meanings of its terms.)

- Frege, G.** (1903, 1906) 'On the Foundations of Geometry', in B. McGuinness (ed.) *Collected Papers on Mathematical Logic and Philosophy*, Oxford: Blackwell, 1983. (1903, series 1, 1906, series 2. Further articulates Frege's critique of implicit definition. See §3.)
- Horwich, P.** (1986) 'A Defence of Conventionalism', in G. MacDonald and C. Wright (eds) *Fact, Science and Morality*, Oxford: Blackwell. (Argues that Quine's critique leaves open a form of 'global conventionalism' concerning the language with which to express our total theory formulation. See §6.)
- Horwich, P.** (1997) 'Implicit Definition, Analytic Truth, and A Priori Knowledge', *Nous*. (Argues that objections to the standard conception of implicit definition may be dissolved in the context of a use-theory of meaning. See §§3, 4, 6.)
- Lewis, D.** (1970) 'How to Define Theoretical Terms', *Journal of Philosophy* LVII (July): 427-66. (Articulates the strategy, suggested by Carnap, Russell and Ramsey in the 1920s, that any theory can be factored into two independent components: a substantive existential claim and an a priori conditional which defines the theoretical terms. See §6.)
- Poincaré, H.** (1905) *Science and Hypothesis*, London: Walter Scott Publishing. (Maintains that the facts of geometry are mere matters of convention. See introduction and §1.)
- Prior, A.N.** (1960) 'The Runabout Inference-Ticket', *Analysis* 21, 38-9. (Shows that an alleged implicit definition presupposes what may well be false: that there exists a meaning that, if possessed by the term in question, would render true the sentences defining the term. See §3.)
- Quine, W.V.** (1935) 'Truth by Convention', in *The Ways of Paradox*, Cambridge, MA: Harvard University Press, 1976. (Argues that logic cannot derive from explicit conventions. See introduction and §2.)
- Quine, W.V.** (1951) 'Two Dogmas of Empiricism', in *From a Logical Point of View*, Cambridge, MA: Harvard University Press, 1953. (Suggests that the alleged distinction between analytic and synthetic statements (and related notions such as synonymy and rule of language) resist clarification and are of dubious coherence. See introduction and §§2, 4.)
- Quine, W.V.** (1954) 'Carnap and Logical Truth', in *The Ways of Paradox*, Cambridge, MA: Harvard University Press, 1976. (Argues against the view that the fundamental principles of logic are conventions of language. See introduction and §§2, 4.)
- Reichenbach, H.** (1968) *The Philosophy of Space and Time*, New York: Dover. (Maintains that assumptions about the physical manifestation of straight lines and other geometrical objects are matters of convention, hence that geometry, which results from such assumptions, is also conventional. See introduction and §1.)

Counterfactual conditionals

'If bats were deaf, they would hunt during the day.' What you have just read is called a 'counterfactual' conditional; it is an 'If...then...' statement the components of which are 'counter to fact', in this case counter to the fact that bats hear well and sleep during the day. Among the analyses proposed for such statements, two have been especially prominent. According to the first, a counterfactual asserts that there is a sound argument from the antecedent ('bats are deaf') to the consequent ('bats hunt during the day'). The argument uses certain implicit background conditions and laws of nature as additional premises. A variant of this analysis says that a counterfactual is itself a condensed version of such an argument. The analysis is called 'metalinguistic' because of its reference to linguistic items such as premises and arguments. The second analysis refers instead to possible worlds. (One may think of possible worlds as ways things might have gone.) This analysis says that the example is true just in case bats hunt during the day in the closest possible world(s) where they are deaf

1 Classification issues and familiar logical connectives

'Counterfactual' conditionals serve to summarize hypothetical reasoning and to justify or explain action and inaction ('I didn't tell you because, had I told you, everyone in the office would have known'). They serve to state particular causal contingencies ('If the battery were flat, the starter would not turn') and various kinds of necessity ('If he were dead, she would be a widow'). Counterfactuals have also been enlisted for philosophical analyses of dispositions ('X is soluble just in case it would dissolve if it were immersed in water'), causality ('Had A not occurred, B would not have occurred'), freedom of action ('I would have acted otherwise if I had so chosen'), knowledge ('If p were false, Mary would not believe that p') and laws of nature, which have been characterized as generalizations that are not merely true but support counterfactuals.

Counterfactuals are commonly contrasted with indicative conditionals (see [Indicative conditionals](#)). The following pair conveys a feeling for the difference:

- (1) If Shakespeare hadn't written *Hamlet*, then someone else would have (counterfactual).
- (2) If Shakespeare didn't write *Hamlet*, then someone else did (indicative).

Instead of 'counterfactual conditional', sometimes the term 'subjunctive conditional' is used. The former refers to a property of the statement made by uttering a conditional sentence in an appropriate context, the latter refers to the grammatical form of the sentence uttered. Neither label is satisfactory. A problem with 'counterfactual' is that not all assertions of counterfactuals are, or are meant to be, counter to fact, as, for example, 'If you did the cooking, I would clean up afterwards' (put forward as a suggestion before the cooking has started).

'Subjunctive' is problematic because the grammatical form in question is not the subjunctive in English (where this form is all but extinct) nor in other European languages (where it is widely used). Moreover, 'subjunctive' contrasts with 'indicative', yet indicatives such as (3) below go much more naturally with 'subjunctives' such as (1) than with indicatives such as (2):

- (3) If Shakespeare does not write *Hamlet*, then someone else will.

It is surprisingly difficult to characterize the difference between the two kinds of conditional. It is not even agreed at which linguistic level - morphosyntactic, semantic or pragmatic - the distinction is to be drawn. We will content ourselves here with defining counterfactual *sentences* as the likes of (1) above. Our question is then which *statements* these sentences normally serve to express.

In the view of most authors, an analysis must state non-circular and informative truth-conditions. It should also be 'compositional': the truth-conditions of ' $A \rightarrow B$ ' should be a function of the truth-conditions of A and B (see [Compositionality](#)). (Read ' $A \rightarrow B$ ' as 'If it were the case that A, then it would be the case that B'.) The material conditional of classical logic and the strict conditional of modal logic both meet these demands, but their truth-conditions are unsuitable. A material conditional is true just in case either its antecedent is false or its consequent true. Since most counterfactuals have false antecedents, they would come out true irrespective of their consequents if they were treated as material conditionals. But many are not true (for example, (1) above); hence the material conditional cannot be the right analysis. A strict conditional is true just in case the corresponding

material conditional is necessarily true, or, in possible worlds parlance, just in case in all worlds in which the antecedent is true the consequent is also true. If the material conditional is too easy to satisfy, the strict conditional is too hard. Consider

(4) If fewer people smoked, fewer would contract lung cancer,

which is true if any counterfactual is. Yet there are possible worlds where the antecedent is true and the consequent false, for example, worlds in which lung cancer is caused by eating chocolate. So the strict conditional is too strict. We might confine our attention to less exotic worlds; ones governed by the same natural laws as the actual world. But that still would not do because among the lawful worlds where fewer people smoke there are worlds where more people work in coal mines and the overall incidence of lung cancer is the same. The worlds in which we are interested must resemble the actual world more closely. We imagine that fewer people smoke in these worlds, but that otherwise things are very much as in the actual world. It is this idea of a minimal departure from actuality (or purported actuality) that underlies, in one form or another, all theories of counterfactuals.

2 The metalinguistic view

Nelson Goodman (1954) put forward an influential view in which ' $A \rightarrow B$ ' is true if and only if (1) there is an appropriate subset of all the true initial conditions that, in conjunction with A and the laws of nature, entails B ; and (2) there is no other subset of the initial conditions that in the same manner entails not- B . The minimal departure from actuality lies in the appropriate choice of initial conditions. Two problems beset Goodman's proposal: the characterization of the relevant initial conditions and the definition of laws. The second problem is serious, but not as serious as the first.

Apart from being true, the relevant initial conditions must be jointly compatible with A (as long as A is consistent), for otherwise any consequent whatever could be inferred and the corresponding counterfactual would be true. The conditions must also be compatible with not- B , for otherwise A and the laws would play no role in the inference to B - B would follow from the initial conditions by themselves. ('Compatible' statements are ones which do not jointly entail a contradiction.) But compatibility alone is not a strong enough condition. Consider

(5) If match m had been struck, it would have lighted.

Let O = 'oxygen is present', D = ' m is dry', S = ' m is struck', L = ' m lights' and N = 'If O , D and S , then L ' (our natural law). Suppose that we start with the situation of O , D and N being true and that the match has not been struck and has not lighted. O , D and N are jointly compatible with S and jointly compatible with not- L , and together with S they imply L . Therefore (5) should be true.

However, not- L is also true, not- L , O and N are jointly compatible with S and jointly compatible with D ; and O , N and not- L together with S imply not- D . Thus

(6) If match m had been struck, it would not have been dry

should be true as well. This is unacceptable. Hence logical compatibility is not enough. The relevant conditions must be jointly tenable or 'cotenable' with the antecedent. (B is said to be cotenable with A if it is not the case that B would be false if A were true.) In this case, not lighting is compatible with, but not cotenable with, striking, because, had m been struck, it would have lighted; so (5) is true and (6) is not. But now we are caught in a circle or regress because cotenability is defined in terms of counterfactuals, while counterfactuals are defined in terms of cotenability. Goodman saw no way out of this quandary.

3 Possible worlds semantics

In the 1960s, modal logicians turned talk of possible worlds into a powerful analytic tool. Robert Stalnaker and David Lewis applied it to the analysis of counterfactuals. Both saw counterfactuals as devices to locate the actual world within a certain similarity structure in logical space. The intuition they built on is that ' $A \rightarrow B$ ' is true in the actual world just in case B is true in the most similar A -world(s) (an A -world being a world where A is true). Stalnaker (1968) represents the minimal departure from actuality by a selection function on possible worlds, Lewis (1973) by ordering worlds so that given any two worlds u and v , either u is more similar to the actual world than v , or vice versa, or else u and v are equally similar to the actual world. For Stalnaker, ' $A \rightarrow B$ ' is true if and only if

B is true at $f(A, w^*)$, the world picked out by applying the selection function f to antecedent A and the actual world w^* . f selects a world in which A is true and which, intuitively, resembles w^* as closely as the truth of A permits. For Lewis, ' $A \rightarrow B$ ' is true if and only if either there is no A -world, or some A & B -world is more similar to the actual world than any A & not- B -world. Lewis' analysis makes ' $A \rightarrow B$ ' trivially true when A is impossible, which is when there is no A -world. So does Stalnaker's analysis because, for impossible A , f selects 'the impossible world' in which every statement is true.

An argument form is valid if and only if it has no instance with true premise(s) and a false conclusion. The following is an example of an invalid argument:

If Ronald Reagan had been born a Russian, he would have been a communist.
 If he had been a communist, he would have been a traitor.
 Therefore, if he had been born a Russian, he would have been a traitor.

The premises seem true, but not the conclusion. The argument is an instance of 'hypothetical syllogism':

$A \rightarrow B, \quad B \rightarrow C \quad \therefore A \rightarrow C$

The possible worlds theory explains how hypothetical syllogism can fail to be valid. World c , the closest world where Reagan was a communist, is closer to the actual world than world r , the closest world where he was born a Russian. In c , his career is as we know it, except for that traitorous affiliation. In r , he grows up in Russia, becomes a loyal party member there and never dreams of being president of the USA. So c makes the second premise of the argument true, while r makes the first premise true and the conclusion false.

On the Lewis-Stalnaker analyses, a number of argument forms that are valid for the material and various strict conditionals are invalid. Hypothetical syllogism is only one of them. Others are:

$A \rightarrow B \therefore \text{not-}B \rightarrow \text{not-}A$ ('contraposition')
 $A \rightarrow B \therefore C \rightarrow B$, where C entails A ('strengthening the antecedent')
 $A \rightarrow (B \rightarrow C) \therefore (A \& B) \rightarrow C$ and
 $(A \& B) \rightarrow C \therefore A \rightarrow (B \rightarrow C)$ ('import-export')

For practical purposes, the invalid argument forms can often be replaced by similar valid ones, for example, hypothetical syllogism by

$A \rightarrow B, \quad (A \& B) \rightarrow C \quad \therefore A \rightarrow C.$

The distinction between valid and invalid argument forms ought to match natural language usage. Examples like the one about Reagan therefore carry heavy weight. If they were rejected and any of the argument forms in question added to the logics that emerge from the Lewis-Stalnaker analyses, ' $\rightarrow$ ' would collapse into the material conditional. Attacks on the counterexamples are therefore indirect attacks on the core of the analyses. (For more discussion, see Stalnaker 1984.)

4 Similarity

Lewis allows ties in the relative similarity between worlds, as well as no limit to how closely worlds can resemble each other. Stalnaker's theory can be seen as a special case of Lewis' that does not make these allowances. For Stalnaker, there is to be always exactly one most similar antecedent-world. The difference is reflected in a disagreement about the principle of 'conditional excluded middle' which says that ' $A \rightarrow B$ ' and ' $A \rightarrow \text{not-}B$ ' cannot both be false. The principle holds in Stalnaker's view but not in Lewis'. In Lewis' view, there may be, for any A & B -world, an equidistant or closer A & not- B -world and vice versa, so that both counterfactuals can be false. For Stalnaker, there is exactly one closest A -world in which either B or not- B is true. Consider the pair

If Bizet and Verdi had been compatriots, Bizet would have been Italian.
 If Bizet and Verdi had been compatriots, Verdi would have been French.

We hesitate to affirm either. Doesn't this show that Lewis is right? Not necessarily. We also hesitate to deny either of the two, and the best explanation may be that the relevant standard of similarity between worlds has not been

made clear, so that we do not know exactly what claims are at issue. Conditional excluded middle holds relative to a given selection function f , but it does not apply to cases where f has not been determined. Any theory of counterfactuals must allow for a good measure of indeterminacy if it is to account for actual usage.

Goodman's theory can be transposed into the possible worlds framework. With the right definition of cotenability, it becomes equivalent to Lewis'. Does Lewis therefore solve Goodman's problem about cotenability? In some sense, yes, because there is no longer a circle in the truth-conditional schema. In another sense, no, because Goodman's worry was less about the abstract form of the analysis than about how particular cotenability judgments can be grounded. For an answer to this question, the notion of relative similarity between worlds is probably too vague.

The relevant notion of similarity cannot be the one we use when we compare ordinary objects. For otherwise we should reject

If Nixon had pushed the button, there would have been a nuclear holocaust,

because among the worlds where Nixon pushes the ominous button those where nothing happens are more similar to ours, in the everyday sense of similarity, than those where the doomsday machinery goes off. But given that there is (or was) such a button, the counterfactual looks true. Lewis suggests that the similarity relation relevant for counterfactuals has the most similar worlds agree with the actual world throughout the past until shortly before the time of the antecedent. Then a small miracle (a deviation from the laws of the actual world, not from those of the world in question) brings about the antecedent, and from then on history evolves in accordance with the laws of the actual world. The proposal has limited scope. It does not apply to 'counterlegals' about what would be the case if this or that law of nature were different. It also does not deal with disjunctive antecedents pertaining to different times. Maximizing agreement throughout the past would naturally lead here to a selection of worlds in which only the disjunct pertaining to the latest time holds. Other authors favour total agreement in laws over miracles of any size. Miracles prevent a counterfactual extrapolation of what the past would have had to have been in order to bring about the antecedent. Yet such 'backtracking' is perfectly intelligible. Jim and Jack quarrelled yesterday and Jack is still furious. If Jim asked a favour of Jack today, Jack would oblige him none the less. For in order that Jim ask, there would have had to have been no quarrel before. The significance of backtracking and Lewis' proposal about similarity remain contested.

5 Comparisons

Metaphysical and other qualms about the possible worlds analyses rekindled an interest in metalinguistic theories in the 1980s. There have also been proposals for an epistemic analysis based on the idea that ' $A \rightarrow B$ ' is 'acceptable' with respect to a body of beliefs K just in case K , minimally changed so as to accommodate A , entails B . Formally, most of these theories are equivalent to Lewis' or slight variants thereof, but they avoid the ontological cost of possible worlds. In so far as they do not provide truth-conditions, the epistemic theories have difficulty explaining the role of counterfactuals embedded in truth-conditional constructions such as conjunctions or disjunctions. Moreover, the idea of revising a body of beliefs seems better suited for evaluating indicative conditionals than counterfactuals. For example, if you add to your present beliefs the assumption that Shakespeare did not write *Hamlet* and make the minimal changes necessary to restore consistency, you will infer that someone else must have written the piece because you have no reason to give up your belief in its existence. This yields the correct judgment for the indicative (2) but not for the counterfactual (1). It is not obvious how to adapt procedures of belief change to counterfactuals.

Goodman and many others believed that an analysis of counterfactuals was indispensable for the philosophy of science. Few still share this opinion. The cotenability problem, the indeterminacy of many counterfactuals, and doubts about possible worlds and their similarities have convinced many that counterfactuals do not belong in the ultimate scientific description of the world. In decision and game theory, by contrast, the peculiarity and importance of counterfactual reasoning has only just begun to be appreciated. The solution of decision problems requires answers to questions about situations that, as a result of the very decision at issue, will be counterfactual. Counterfactuals also continue to play an important role in philosophical analyses, and their use and misuse in daily commerce is largely unhampered by difficulties of interpretation.

See also: [Logical and mathematical terms, glossary of](#); [Possible worlds](#); [Relevance logic and entailment](#)

References and further reading

- Edgington, D.** (1995) 'On Conditionals', *Mind* 104.414: 235-329. (Gives an excellent overview and synthesis of the various issues surrounding conditionals.)
- Goodman, N.** (1954) *Fact, Fiction, and Forecast*, Atlantic Highlands, NJ: Athlone Press. (See §3. A non-technical classic.)
- Harper, W.L., Stalnaker, R. and Pearce, G.** (eds) (1981) *IFS: Conditionals, Belief, Decision, Chance, and Time*, Dordrecht and Boston, MA: Reidel. (A collection of important papers of varying degrees of technicality.)
- Jackson, F.** (ed.) (1991) *Conditionals: Oxford Readings in Philosophy*, Oxford: Oxford University Press. (A collection of important papers, including Lewis' on the similarity between worlds.)
- Lewis, D.K.** (1973) *Counterfactuals*, Cambridge, MA: Harvard University Press. (See §4. The philosophical discussion demands no technical expertise, but the details of the semantics and the logic get very technical.)
- Sanford, D.H.** (1989) *If P then Q: Conditionals and the Foundations of Reasoning*, London and New York: Routledge. (An informative historical overview.)
- Stalnaker, R.** (1968) 'A Theory of Conditionals', *Studies in Logical Theory: American Philosophical Quarterly Monograph Series* 2: 98-112; repr. in W.L. Harper, R. Stalnaker and G. Pearce (eds) *IFS: Conditionals, Belief, Decision, Chance, and Time*, Dordrecht and Boston, MA: Reidel, 1981, 41-56. (See §4. The first statement of Stalnaker's position; no technical expertise required.)
- Stalnaker, R.** (1984) *Inquiry*, Cambridge, MA: Bradford Books. (An excellent non-technical discussion of counterfactual and indicative conditionals and their role in reasoning.)

De re/de dicto

'De re' and 'de dicto' have been used to label a host of different, albeit interrelated, distinctions. 'De dicto' means 'of, or concerning, a dictum', that is, something having representative content, such as a sentence, statement or proposition. 'De re' means 'of, or concerning, a thing'. For example, a *de dicto* belief is a belief that a bearer of representative content is true, while a *de re* belief is a belief concerning some thing, that it has a particular characteristic.

Consider the following example:

John believes his next-door neighbour is a Buddhist.

This statement is ambiguous. Construed *de dicto*, it is true in the following circumstance. John has never had any contact with his next-door neighbour. Nevertheless, John believes that his next-door neighbour is bound to be a Buddhist. Construed in this *de dicto* fashion, the statement does not attribute to John a belief that is distinctively about a particular individual. In contrast, construed *de re*, it does attribute to John a belief that is about a particular individual. For example, construed *de re*, the statement is true in the following circumstance. John encounters his next-door neighbour, Fred, at a party without realizing that Fred is his next-door neighbour. On the basis of his conversation with Fred, John forms a belief about the individual who is in fact his next-door neighbour to the effect that he is a Buddhist.

1 The different distinctions

'De dicto' means 'of, or concerning, a dictum'. 'De re' means 'of, or concerning, a thing'. Roughly, a *de dicto* belief is a belief that a bearer of representative content (that is, a statement, sentence, proposition or such like) is true, while a *de re* belief is a belief concerning some thing, that it has a particular characteristic. (I)-(IX) below are a number of different distinctions that have been described as *de re/de dicto* distinctions in the literature.

(I) In the case of modality and belief we can draw scope distinctions. Consider the following sentences:

- (1) Every effect necessarily has a cause.
- (2) Joan believes her husband is an axe murderer.

(1) and (2) manifest a scope ambiguity. (1) is ambiguous between:

- (1') [squ] (if *e* is an effect then *e* has a cause),

in which the entire sentence 'If *e* is an effect then *e* has a cause' falls within the scope of 'Čsqu;'; and:

- (1*) If *e* is an effect, [squ] (*e* has a cause),

in which only the part '*e* has a cause' falls within the scope of 'Čsqu;'. (2) may, likewise, be disambiguated as:

- (2') Joan believes that her husband is an axe murderer

or:

- (2*) Joan's husband is believed by Joan to be an axe murderer.

(1') and (2') give the *de dicto* readings of (1) and (2). (1*) and (2*) give their *de re* readings. According to (1'), it is impossible for something to be an effect without having a cause - something few would dispute. According to (1*), it is impossible for anything which is an effect not to have been caused - a much more controversial claim. (2') implies that Joan has the disconcerting belief she could express by uttering, 'My husband is an axe murderer'. (2*) implies that Joan has a less disconcerting belief. According to (2*), she believes that someone, whom, perhaps unbeknownst to her, she is married to, is an axe murderer (see [Propositional attitudes §2](#)).

(II) A sentence expresses a *de re* modality provided that some object essentially features in its truth-condition in the following sense: its truth-value depends on the characteristics that the object might have. For example, the

sentence ‘Possibly Fred is a Buddhist’ expresses a *de re* modality because its truth-value depends on whether a certain individual, Fred, could have the characteristic of being a Buddhist.

(III) Some hold that predicates such as ‘is necessarily human’ or ‘is possibly six feet tall’ attribute a special kind of property known as a modal property. A *de re* modal attribution attributes a modal property. For example, ‘John is necessarily human’ expresses a *de re* modal attribution if it attributes to John the modal property of being necessarily human.

(IV) Another *de re/de dicto* distinction depends on the distinction between rigid and non-rigid designators. Proper names such as ‘Napoleon’ are said to be rigid designators because they designate the same thing in any possible situation in which they designate at all (see **Proper names §2**). According to this way of drawing the distinction, a modal sentence expresses a *de re* modality if it contains a rigid designator falling within the scope of a modal operator. For example, ‘It might have been the case that Napoleon spoke Chinese’ expresses a *de re* modality, in contrast to ‘It might have been the case that the most famous French general spoke Chinese’.

(V) Quine (1953) insists that modalities are invariably relativized to a description in the following sense. We cannot say that something necessarily or possibly has a characteristic absolutely: we can only say that something necessarily or possibly has a characteristic relative to a description. Bill Clinton is necessarily a president relative to the description ‘42nd President’. He is contingently a president relative to the description ‘husband of Hillary’. In Quine’s view all modalities are *de dicto*. *De re* modalities, if, contra Quine, they existed, would be those not relativized to a description.

Some further distinctions taken to be distinctions between the *de re* and the *de dicto* are illustrated in application to belief.

(VI) It has been argued that we should distinguish between two kinds of belief. Having a *de dicto* belief amounts to standing in a two place relation to a dictum. Having a *de re* belief amounts to standing in an at least three place relation to an object and a property. Suppose John believes that all prominent politicians are wealthy. In consequence, without knowing who is president, he believes that the president is wealthy. In that case, John will have a *de dicto* belief in virtue of standing in the two place belief relation to the sentence ‘The president is wealthy’.

Now suppose that Sally meets Bill Clinton, and judges from the cut of his clothes that he is wealthy. Whether or not she believes he is president, Sally believes of the president that he is wealthy. According to this way of distinguishing *de re* from *de dicto* beliefs, Sally stands in the three place belief relation to the president and the characteristic of being wealthy.

(VII) In another way of distinguishing between *de re* and *de dicto* beliefs, a *de re* belief is about a particular individual or individuals, whereas a *de dicto* belief is general. Sally’s belief about the president that he is wealthy is a belief about a particular individual to the effect that he is wealthy. On the other hand, John’s belief that the president is wealthy is the general belief that whoever occupies the presidency is wealthy.

(VIII) A different epistemic distinction between *de re* and *de dicto* beliefs is as follows. In contrast with a *de dicto* belief, in order to have a *de re* belief one must be epistemically acquainted with the subject of one’s belief: one must know who, which or what it is.

(IX) A final distinction is that between object-dependent and object-independent beliefs. Sally sees a cat and forms the belief that she expresses by uttering, ‘That cat has mange’. Some say that she cannot hold this belief about the cat unless it exists. If so, her belief is object-dependent. It is sometimes held that *de re*, as opposed to *de dicto*, beliefs are object-dependent in this sense.

2 Relationships between the different distinctions

Consider the following examples:

(3) [squ] (the subject of Gance’s best-known film was human).

(3’) The subject of Gance’s best-known film is such that [squ] it was human.

Abel Gance’s best-known film was about Napoleon. According to distinction (I) above, sentence (3) expresses a

de dicto modality and (3') a *de re* modality. 'The subject of Gance's best-known film' falls within the scope of 'Čsqu;' in (3), and outside the scope of that modal operator in (3'). So Napoleon does not essentially feature in the truth-condition of (3). Hence, distinctions (I) and (II) above would agree in counting (3) as a *de dicto* modality. Since 'the subject of Gance's best-known film' falls outside the scope of 'Čsqu;' in (3'), (3') is true if and only if it is impossible for Napoleon to have been nonhuman. So (I) and (II) likewise agree in counting (3') as a *de re* modality.

Distinction (III) has it that sentences expressing *de re* modalities attribute modal properties. The truth-value of (3) does not depend on Napoleon having the modal property of being necessarily human. Arguably, the truth-value of (3') does depend on his having that modal property. So (III), like (I) and (II), decrees that (3) is a *de dicto* modality and (3') is *de re*.

'The subject of Gance's best-known film' is a non-rigid designator because it designates different things in different possible situations. Since 'the subject of Gance's best-known film' falls within the scope of the modal operator in (3), (IV) deems (3) to be a *de dicto* modality. In (3'), 'the subject of Gance's best-known film' falling outside the scope of 'Čsqu;' has the same effect as replacing 'the subject of Gance's best-known film' with a rigid designator such as 'Napoleon' in (3).

(3*) [squ] (Napoleon was human).

The truth-value of (3*) depends on whether the actual referent of 'Napoleon' might have been nonhuman. According to (IV), (3*) expresses a *de re* modality for basically the same reason that (3') does according to (II) and (III).

Distinction (V) implies that a modality is *de dicto* if it is relativized to a description. Otherwise, it is *de re*. (3) is true if and only if nothing could be nonhuman which satisfies the description 'the subject of Gance's best-known film'. In Quine's sense the modality in (3) is relativized to a description, and is an example of a *de dicto* modality. In contrast, since the definite description in (3') falls outside the scope of the modal operator, the truth of (3') requires that the individual who actually satisfies that definite description must be human. (3') is true if and only if Napoleon, the individual actually referred to by 'the subject of Gance's best-known film', must be human. (V) implies that (3') is a *de re* modality.

Now consider:

(4) John believes that the president is wealthy.

(4') The president is such that Sally believes he is wealthy.

According to (VI) above, a *de dicto* belief relates a believer to a proposition, sentence or statement, whereas a *de re* belief relates a believer to something that may be extralinguistic. John's *de dicto* belief that the president is wealthy relates John to the sentence 'The president is wealthy'. The *de re* belief attributed to Sally by (4') relates her to Bill Clinton.

Since Sally's *de re* belief about the president relates her to Clinton, she could not have that belief unless Clinton exists. Hence, the belief attributed to Sally by (4') turns out to be object-dependent in conformity with distinction (IX). Moreover, if Sally's *de re* belief about the president relates her to Clinton, it is about a specific individual. Hence, it is *de re* in conformity with (VII). Since Sally knows who it is she believes to be wealthy, her belief is also *de re* in the terms of distinction (VIII).

See also: [Essentialism](#); [Logical and mathematical terms, glossary of](#); [Use/mention distinction and quotation](#)

ANDRÉ GALLOIS

References and further reading

- Burge, T.** (1977) 'Belief *De Re*', *Journal of Philosophy* 74: 317-38. (A classic attempt to articulate the notion of a *de re* belief, and argue that it is more fundamental than the notion of a *de dicto* belief.)
- Dennett, D.** (1982) 'Beyond Belief', in A. Woodfield (ed.) *Thought and Object*, Oxford and New York: Oxford University Press. (A sustained attempt to call into question the distinction between *de re* and *de dicto* beliefs.)
- Fine, K.** (1989) 'The Problem of *De Re* Modality', in J. Almog, J. Perry and H.K. Wettstein (eds) *Themes From*

- Kaplan*, Oxford and New York: Oxford University Press, 197-272.(A somewhat difficult, but rewarding, discussion of key distinctions related to *de re* modalities.)
- Kneale, W.** (1960) 'Modality *De Dicto* and *De Re*', in E. Nagel, P. Suppes and A. Tarski (eds) *Logic, Methodology and the Philosophy of Science: Proceedings of the 1960 International Conference*, Stanford, CA: Stanford University Press, 1962, 622-33.(Particularly useful in providing a historical background to the distinction between *de re* and *de dicto* modalities.)
- Kraut, R.** (1983) 'There are No *De Dicto* Attitudes', *Synthese* 54: 275-94.(A defence of scepticism about the distinction between *de re* and *de dicto* attitudes.)
- McDowell, J.** (1984) '*De Re* Senses', *Philosophical Quarterly* 34: 283-94.(Includes a discussion of the alleged object dependence of *de re* thoughts.)
- Plantinga, A.** (1969) '*De dicto* et *de re*', *Noûs* (3): 235-58.(An attempt to reduce *de re* to *de dicto* modalities.)
- Quine, W.V.** (1953) 'Reference and Modality', in *From a Logical Point of View*, New York: Harper, 139-59. (Includes Quine's argument against the meaningfulness of *de re* modal attributions.)
- Quine, W.V.** (1976) 'Quantifiers and Propositional Attitudes', in *Ways of Paradox and Other Essays*, Cambridge: Cambridge University Press, 185-96.(A classic statement of one way of distinguishing between *de re* and *de dicto* beliefs.)
- Quine, W.V.** (1979) 'Intensions Revisited', in P.A. French, T.E. Uehling, Jr and H.K. Wettstein (eds) *Contemporary Perspectives in the Philosophy of Language*, Minneapolis, MN: University of Minnesota Press, 268-74.(Quine's second thoughts about the *de re/de dicto* distinction.)

Decision and game theory

Decision theory studies individual decision-making in situations in which an individual's choice neither affects nor is affected by other individuals' choices; while game theory studies decision-making in situations where individuals' choices do affect each other. Decision theory asks questions like: what does it mean to choose rationally? How should we make choices when the consequences of our actions are uncertain? Buying insurance and deciding which job to take are examples of the kind of decisions studied by this discipline. Game theory instead applies to all decisions that have a strategic component. The choices of an oligopolist, voting strategies, military tactical problems, deterrence, but also common phenomena such as threatening, promising, conflict and cooperation are its subject matter. In a strategic situation, the goal is not just to choose rationally, but to choose in such a way that a mutual solution is achieved, so that choices 'coordinate' in the right way. The formal methods developed by game theory do not require that the subject making a choice be an intentional agent: coordinated interaction between animals or computers can be successfully modelled as well.

1 Probability as a guide to life

The thesis that reasonable decisions must be based on probability was already regarded as common-sensical in the eighteenth century. It is presented most clearly in Joseph Butler (1736), who claimed that 'probability is the very guide to life'. The thesis was first explicitly formulated by Antoine Arnauld (1662), who aimed to dispel the common illusion 'of considering only the magnitude and importance of the expected gain or the feared loss rather than considering the probability of the gain or loss'. A decision made under conditions of uncertainty thus had to depend upon two factors: an evaluation of the probability of the possible events and an evaluation of the gains or losses following such events. Take, for example, a gamble in which one of n outcomes will occur. Let the outcomes be worth $x_1, x_2, \dots, x_n$ dollars, respectively. The probability of each outcome is known, therefore we have $p_1, p_2, \dots, p_n$, where $0 \leq p \leq 1$ and $\sum_{i=1}^n p_i = 1$. How much is it worth to participate in this gamble? The answer was that the expected monetary value $\sum_{i=1}^n x_i p_i$ is the fair 'price' of the gamble.

For centuries, then, it was taken for granted that proper choice meant maximizing expected monetary value. Suppose, however, that you face the following gamble. A fair coin is to be tossed until a head appears. You receive 2^n dollars if the first head occurs on the n th trial. The expected monetary value of this gamble is $\sum_{n=1}^{\infty} 2^n (1/2)^n = 1 + 1 + 1 + \dots$, an infinite value. Thus, on the account we are considering, you should be willing to pay any finite stake to play it. Introspection tells us that no one will act like that: how then do we explain the discrepancy between theory and experience?

In 1738 Daniel Bernoulli, provoked by the above paradox (the so-called St Petersburg paradox), suggested that to determine the value of a gamble one should not consider its monetary value but instead the *utility* that such money has for the gambler. Bernoulli recognized that the utility of a given sum of money may not be the same for different people, and may even differ for the same person at different times. In general, any increase in wealth will produce a utility increase that is inversely proportional to the quantity of goods already possessed. According to Bernoulli, one should maximize some concave function of money (he suggested the logarithmic function), to reflect its diminishing marginal utility. In the example above, if we take the utility of a sum of money to be proportional to its logarithm, and we take the mathematical expectation of utility (rather than money), we have a finite value, so that the gambler is willing to pay only a finite stake. The fair price of the gamble is now the monetary equivalent of its expected utility value, for example, $\sum_{n=1}^{\infty} \log_{10} 2^n (1/2)^n = \log_{10} 2 \times \sum_{n=1}^{\infty} n(1/2^n)$. In the limit, this sum approaches a finite value. Expected utility was born.

2 Expected utility

Almost 200 years later, von Neumann and Morgenstern (1944) proved that if individual preferences satisfy a given set of axioms, they can be represented by the expectation of some utility function. Their theory applies to agents who must choose one among a set of feasible actions without being able to predict with certainty what the outcome of a given choice will be. Each action may result in one of several possible outcomes, depending on which state of the world occurs, and there exists a known probability distribution over the states. In this case, we say that the

choice is one under *risk*, to distinguish it from a choice situation in which there are no known probabilities. Their results initiated a vast research field known as *decision theory*. We shall highlight here only a few of the main concepts of decision theory that are particularly relevant to game theory.

Suppose action A has two possible outcomes, a and a' , and that these outcomes occur with probabilities p and $(1 - p)$, respectively. Choosing A is thus like choosing a lottery that gives you prize a with probability p and prize a' with probability $(1 - p)$. Denote this lottery by $(ap, a'(1 - p))$. Agents are completely defined by their preferences over such lotteries. Let us denote lotteries by X, Y, Z , and probabilistic mixtures of lotteries (compound lotteries) by $pX + (1 - p)Y$. Specific outcomes are just degenerate lotteries with an outcome probability of one, and are denoted by x, y, z . Denote ' X preferred to Y ' by $X \succ Y$, ' X indifferent to Y ' by $X \sim Y$. The von Neumann-Morgenstern axioms concerning individuals' preferences over lotteries are:

- (1) *Ordering*. Preferences are complete: either $X \succ Y$, $Y \succ X$, or $X \sim Y$, and transitive: $X \succ Y$ and $Y \succ Z$ implies $X \succ Z$.
- (2) *Preference increasing with probability*. If $x \succ x'$ and $X = (px, (1 - p)x')$, $Y = (qx, (1 - q)x')$, then $X \succ Y$ if and only if $p > q$.
- (3) *Continuity*. For all lotteries $X \succ Y \succ Z$, there exists a unique p such that $(pX, (1 - p)Z) \sim Y$. (A consequence of this axiom is the following: given that there is a best lottery outcome b and a worst lottery outcome w , for any lottery outcome x there will always be some probability combination of b and w such that the agent is indifferent between that combination and x . For given x , define u_x such that $x \sim (u_x b, (1 - u_x)w)$. u_x is called the von Neumann-Morgenstern utility index of x .)
- (4) *Independence*. If $X \succ Y$, then $(pX, (1 - p)Z) \succ (pY, (1 - p)Z)$, for all $p \in (0, 1)$ and Z .
- (5) *Rules for combining probabilities*. Consider the lottery $(px, (1 - p)x')$. By continuity, we have that $X \sim [p(u_x b, (1 - u_x)w), (1 - p)(u_{x'} b, (1 - u_{x'})w)]$. Denote the compound lottery by X^* . The probability of obtaining b in the lottery X^* is $pu_x + (1 - p)u_{x'}$. Similarly, the probability of obtaining w is $p(1 - u_x) + (1 - p)(1 - u_{x'})$. So X is indifferent to $[pu_x(b) + (1 - p)u_{x'}(b), p(1 - u_x)(w) + (1 - p)(1 - u_{x'})(w)]$. Now consider a lottery $Z = (zq, z'(1 - q))$. It must be the case that $Z \sim [(qu_z(b) + (1 - q)u_{z'}(b), q(1 - u_z)(w) + (1 - q)(1 - u_{z'})(w)]$. Using (2) we have that $X \succ Z$ when $pu_x + (1 - p)u_{x'} > qu_z + (1 - q)u_{z'}$. Since u_x is the utility index of x , we can say that X is preferred to Z when the expected utility of X is greater than the expected utility of Z .

A *rational agent* is understood to be one who chooses an action that best fulfils their goals. This is a minimal definition of rationality, since it does not encompass beliefs (beyond consistency of probability assessments) or goals. Given a goal, an agent will usually have several alternative actions that satisfy it, and a preference ranking over them. A rational agent will act according to that ranking and choose one among the maximally preferred alternatives (there may be more than one). If their preference ordering satisfies conditions (1)-(5), the agent is choosing the action with the highest expected utility.

The major result of von Neumann and Morgenstern is the proof that - when conditions (1)-(5) hold - there exists a real-valued utility function U over consequences (which is unique up to a positive linear transformation) such that for any two lotteries X and Y , $X \succ Y$ if and only if $\sum_{x \in X} p(x)U(x) > \sum_{y \in Y} p(y)U(y)$. A rational decision maker is thus represented as acting so as to maximize his expected utility.

Most choices, however, involve uncertain consequences whose probability cannot be objectively quantified. Savage (1954) proved a representation theorem showing that if preferences for actions are transitive and satisfy certain other mathematical conditions, then there is a utility function on consequences (unique up to a linear transformation) and a unique subjective probability distribution on states of the world such that action A is preferred to action B if and only if A has greater expected utility than B (see *Rationality, practical*).

Expected utility theory was never meant to be descriptive of how agents behave under risk or uncertainty. Rationality should be viewed as an idealization, a useful benchmark for measuring how much actual choices deviate from the ideal. Some of the most interesting recent research on decision theory has been motivated by empirical results which reveal systematic violations of the expected utility axioms, as well as by the so-called paradoxes of Allais (1953) and Ellsberg (1961). Alternative utility models have been proposed by Kahneman and Tversky (1979), Machina (1982) and Fishburn (1982) (see *Rational beliefs*).

Strategic interactions. Many of the consequences of our decisions depend not just on our choice and the occurrence of some state of the world, but also on how *other* agents choose. In a *strategic* context, people interact and make decisions on the basis of expectations about other people's behaviour. These are the kind of situations to which *game theory* applies. Von Neumann and Morgenstern introduced their decision theory as a tool for playing a special type of game in which whatever one player gains the other loses. An example of such a *zero-sum game* is matching pennies, where each player privately turns a penny either heads up or tails up. If the choices match, player 1 gives the penny to player 2; if they do not, player 2 gives player 1 the penny. In this game, rational player 1 should maximize his *security level*, where the security level of a strategy (see §3) is defined as the worst that strategy can bring the player; player 2 instead should minimize her *hazard level*, the best that a strategy can bring the opponent. Using pure strategies, neither player can guarantee himself more than - 1. But if each player randomizes between heads up and tails up with probability 1/2, each can guarantee himself an expected payoff of 0. The expected payoff combination (0, 0) is the only one that yields each player his security level, hence it is - in von Neumann's words - *individually rational*. In 1928 von Neumann proved the so-called *Minimax theorem*: every two-person zero-sum game with finitely many pure strategies (see §3) has precisely one individually rational payoff vector. In von Neumann's early work, outcomes were always expressed in money, and it was implicitly assumed that the marginal utility of money was the same for all players. To fend off objections to this assumption, a monetary outcome was replaced by its utility, and it is precisely the formulation of expected utility theory that allowed for a rigorous treatment of the expected payoffs that result from players' randomization over pure strategies (see §3).

Strictly competitive games like the zero-sum example above are rarely encountered in economic and political applications. Much more common are *mixed-motive games*, in which players' interests are only in partial conflict, or *coordination games*, in which players' interests often coincide. Research in game theory in the last half of the twentieth century has been almost exclusively focused upon the latter kind of games.

3 Games in normal form

A *normal form* game is completely defined by three elements: a list of players $i = 1, \dots, n$; for each player i , a finite set of pure strategies S_i ; a payoff function u_i that gives player i 's payoff $u_i(s)$ for each n -tuple of strategies $(s_1, \dots, s_n)$, where $u_i: \prod_{j=1}^n S_j \rightarrow R$. All players other than some given player i are customarily denoted as ' $-i$ '. A player may want to randomize over his pure strategies; a probability distribution over pure strategies is called a *mixed-strategy* and is denoted by σ_i . Each player's randomization is assumed to be statistically independent of that of the opponents, and the payoffs to a mixed-strategies' profile are the expected values of the corresponding pure strategies payoffs. If the players were jointly to choose a random device and condition their strategy choices on the outcome of this joint randomization, then they would play a *correlated strategy*.

The matrix in [Figure 1](#) depicts a normal form game: each player picks a strategy independently and without communicating with the other, and the outcome is the joint result of these two strategies. The game is one of *complete information*, in that the players know the rules of the game and other players' payoffs. It is usually assumed that players have *common knowledge* that both possess this information (Aumann 1976). If the players are allowed to enter into binding agreements before the game is played, we say that the game is *cooperative*. *Noncooperative games* instead make no allowance for the existence of an enforcement mechanism that would make the terms of the agreement binding on the players. If preplay negotiation leads to an agreement, and there is no enforcement mechanism available, we should ask whether the agreement is self-enforcing, by virtue of its being in the best interest of each player to adhere to.

Nash equilibrium. The standard solution concept for noncooperative games is that of Nash equilibrium (Nash 1951). Informally, a Nash equilibrium specifies players' actions and beliefs such that (i) each player's action is optimal given that players' beliefs about other players' choices; (ii) players' beliefs are correct. Thus an outcome that is not a Nash equilibrium requires either that a player chooses a suboptimal strategy, or that some players 'misperceive' the situation.

		2	
		<i>C</i>	<i>D</i>
1	<i>C</i>	2, 2	0, 3
	<i>D</i>	3, 0	1, 1

Figure 1

More formally, a *Nash equilibrium* is a vector of strategies $(\sigma_1^*, \dots, \sigma_n^*)$, one for each of the n players in the game, such that each σ_i^* is optimal given (or is a best reply to) σ_{-i}^* . Note that optimality is only conditional on a fixed σ_{-i} , not on all possible σ_{-i} .

A common interpretation of Nash equilibrium is that of a self-enforcing agreement. That is, were players to agree in preplay negotiation to play a particular strategy combination, they would have an incentive to stick to the agreement only in case the agreed upon combination is a Nash equilibrium. While this is generally true of pure strategy Nash equilibria, it is not true of equilibria in mixed strategies. Consider the game in [Figure 2](#).

		2	
		t_1	t_2
1	s_1	8, 3	3, 8
	s_2	5, 6	7, 2

Figure 2

This game has no Nash equilibrium in pure strategies, but Nash proved that - provided certain restrictions are imposed on strategy sets and payoff functions - a game will have at least an equilibrium in mixed strategies. Nash's result generalizes von Neumann's theorem (1928) that every game with finitely many strategies has an equilibrium in mixed strategies. Suppose A plays $((4/9)s_1, (5/9)s_2)$. Then if B chooses t_1 , her expected utility is $3(4/9) + 6(5/9) = 42/9$. If B chooses t_2 , she nets $8(4/9) + 2(5/9) = 42/9$. So if A randomizes with probabilities $(4/9, 5/9)$, B is indifferent between t_1 and t_2 and can thus randomize between them. Suppose B chooses $((4/7)t_1, (3/7)t_2)$; then A nets $41/7$ if he plays s_1 , and $41/7$ if he plays s_2 . Thus A is indifferent between s_1 and s_2 , and can randomize between them. The combination $((4/9)s_1, (5/9)s_2), ((4/7)t_1, (3/7)t_2)$ is a mixed-strategy Nash equilibrium. In a mixed-strategy equilibrium, the equilibrium strategy of each player makes the other indifferent between the strategies on which he is randomizing; thus if A were to know that B randomizes

with probabilities $(4/7, 3/7)$, any of his strategies (pure or mixed) would be a best reply to B 's choice and conversely, were B to know that A randomizes with probabilities $4/9, 5/9$, any of her strategies would be a best reply. Paradoxically, if players were to agree to play a mixed-strategy equilibrium, they would have no particular incentive to play their equilibrium mixed strategies. A mixed-strategy equilibrium is thus a self-enforcing agreement only in the weak sense that, given the other players' equilibrium behaviour, each player is indifferent between all the strategies taken from the support of his equilibrium strategy.

A better justification of mixed-strategy equilibria comes from a different understanding of mixed strategies. This is the *purification* idea, according to which a player's mixed-strategy consists in uncertainties in the minds of other players as to the real values of his payoffs (Harsanyi 1973). That is, player A does not randomize between his strategies with probabilities $5/9$ and $4/9$; it is just that player B attaches probability $5/9$ to the event that A has payoffs such that he will pick the first strategy and probability $4/9$ to the event that A has different payoffs, in which case he would choose the second strategy. When players are uncertain about the payoffs of other players, the game is one of *incomplete information*. In this case, players cannot predict what would be a best response for the other players, and consequently cannot determine their own optimal behaviour.

Harsanyi (1967) provided a method for transforming games of incomplete information into games of imperfect information, for which optimal choices and equilibrium behaviour are well defined. The basic idea is simple. A player who is uncertain about another player's payoff will be treated as if he were uncertain about the *type* of player he faces. A player's type embodies the player's private information. In our example, it is information about one's payoff function, but it may include beliefs about other players' payoff functions or even about other players' beliefs. Players' types are randomly drawn by nature according to an initial probability distribution. Each player knows his type but cannot observe nature's choice; players know, however, the prior type distribution as well as all the possible types. Furthermore, the types and the prior distribution are assumed to be common knowledge among the players. A player's strategy choice depends on his type. If a player knows the strategies of the opponents (as functions of their types), and once he is informed about his own type, he will calculate the conditional probability of the opponents' types (given his own type), compute his expected utility using this probability and choose a strategy that maximizes his payoff function. A *Bayesian equilibrium* in a game of incomplete information with a finite number of types for each player, a prior type distribution and a set of pure strategies for each type of player, is a Nash equilibrium of the 'expanded game' where each player can be one of several types, each playing a different strategy. Harsanyi (1973) showed that a mixed-strategy equilibrium of a game of complete information can be justified as the limit of pure strategy equilibria of slightly perturbed games of incomplete information. In a Bayesian game, once players' type-contingent strategies have been computed, each player behaves as if he was facing opponents who play mixed strategies. Uncertainty thus arises from the distribution of types rather than from use of randomizing devices on the part of the players.

Correlated equilibrium. In a game of incomplete information, a player receives a 'signal' from nature about his type. Such signals are assumed to be private and independent. When signals are public and are not independent, we have correlation. Correlation may be perfect, as when the players observe the same random variable (for example, a coin toss), or imperfect, as when there is a random variable that can take three values, x, y, z , and player 1 knows that the realization is either x or a member of $\{y, z\}$, whereas player 2 knows that it is either z or a member of $\{x, y\}$. Even with imperfect correlation, players can do much better than playing a Nash equilibrium. Consider the game in Figure 3.

		2	
		<i>D</i>	<i>H</i>
1	<i>D</i>	2, 2	0, 3
	<i>H</i>	3, 0	-1, -1

Figure 3

There are three Nash equilibria in this game: (H, D) , (D, H) and $((1/2)D, (1/2)H; (1/2)D, (1/2)H)$. When multiple equilibria are present, and there is some conflict of interest among the players, it becomes difficult to see how a preplay negotiation process may end in an agreement to play one of the equilibria. Schelling (1960) proposed the notion of *focal point* to explain how, in such games, an equilibrium could be reached. For example, if player 1 is much more senior than player 2, (H, D) may seem the natural solution to both. However, notions like focal point or salience (Lewis 1969) refer to information external to the game, are difficult to formalize, and presuppose that players share the same information. Even in games of *pure coordination* like, for example, the game in Figure 5, where players' interests perfectly coincide, we may expect (a, c) to be played because it is the only Pareto optimal outcome, but such confidence can only be grounded in the assumption that players have common knowledge that Pareto optimality makes it a focal point for both. In games of pure coordination, however, preplay negotiation will easily identify an equilibrium, therefore the possibility of preplay communication among the players would easily solve the multiplicity problem. The game of Figure 3 is a more difficult case. Obviously player 1 prefers (H, D) and 2 prefers (D, H) . (D, D) is not self-enforcing, and the mixed-strategy equilibrium only offers a meagre expected payoff of 1 to each player. Suppose now that the players can jointly watch the toss of a fair coin, and agree to play (D, H) if heads, and (H, D) if tails. This lottery is worth 1.5 to each player, and $(1.5, 1.5)$ is a better outcome than the mixed-strategy equilibrium. The agreement is self-fulfilling: observing heads establishes (D, H) as a focal point. The case of imperfect correlation is similar. Suppose that the two players engage a referee at no cost. The referee will use a random device to select one of the cells according to the following probabilities: $p(D, D) = p(H, D) = p(D, H) = 1/3$, $p(H, H) = 0$. The referee will not inform the players about the chosen cell, but will secretly tell 1 to use the row in which the cell occurs, and 2 to use the column in which the cell occurs. The players agree to do whatever the referee suggests. If they honour the agreement, (D, D) , (D, H) and (H, D) will be played with probability $1/3$ each. This agreement is self-fulfilling and corresponds to the outcome $(1.66, 1.66)$. It is, in Aumann's words, a *correlated equilibrium*, and it can be proved that the set of correlated equilibria always contains the Nash equilibria of the game (Aumann 1974).

Rationalizability. Notice that, in order to justify equilibrium play, we introduce self-enforcing agreements, and the possibility of preplay negotiation that would allow such agreements to take shape. In most games, common knowledge of the structure of the game and of players' rationality are not sufficient to predict that an equilibrium will be played, even if it is the unique equilibrium for the game. The concept of Nash equilibrium embodies a notion of individual rationality, since each player's equilibrium strategy is a best reply to the opponents' strategies, but unfortunately it does not specify how players can arrive to form the beliefs about each other's strategies that support equilibrium play. Beliefs, that is, can be internally consistent but fail to achieve the interpersonal consistency that guarantees that an equilibrium will be reached. Bernheim (1984) and Pearce (1984) have argued that assuming players' rationality (and common knowledge thereof) can only guarantee that a strategy will be *rationalizable*, in the sense of being supported by coherent beliefs about other players' choices and beliefs. But a combination of rationalizable choices need not constitute a Nash equilibrium. Take the game of Figure 2. Player 1

thinks player 2 will choose t_2 because he might think 1 will play s_1 , on the assumption that 2 will play t_1 because 1 plays s_2 . If 1 is expected to play s_2 , it is rationalizable that 2 plays t_1 . t_1 is a rationalizable strategy because player 2 can tell a coherent story about why 1 might choose s_2 . Note that in this game all four combinations of strategies are rationalizable, yet none of them is an equilibrium.

Dominance. In a Nash equilibrium, the optimality of a strategy is only conditional on a fixed σ_{-i} , not on all possible σ_{-i} . In the game of **Figure 1**, however, the two Nash equilibrium strategies (D, D) are also optimal with respect to any strategy choice of the opponent. Whatever player 1 does, player 2 is better off by choosing D , and the same is true for player 1. We say that D is a *strictly dominant* strategy, in that it guarantees a player a better payoff than any other strategy, irrespective of the opponent's choice. If instead a strategy guarantees a player payoffs that are greater than or equal to the payoffs of any other strategy, that strategy is only *weakly dominant* for the player. Note that a *strictly dominant equilibrium* does not guarantee an optimal outcome; **Figure 1** is the well-known prisoner's dilemma, where (D, D) is the unique, strictly dominant equilibrium, but playing (C, C) would have been better for both players. In this case, rationality dictates choosing (D, D) , and indeed this is the unique prediction for the game. Notice that a strictly dominated strategy is never rationalizable, hence in this game the unique rationalizable combination of strategies is the Nash equilibrium.

Sometimes games can be solved by *iterated dominance*. Consider the game of **Figure 4**.

		2	
		<i>L</i>	<i>R</i>
1	<i>U</i>	8, 10	-100, 9
	<i>D</i>	7, 6	6, 5

Figure 4

L is a strictly dominant strategy for player 2, and since rationality is common knowledge, 2 is expected to eliminate R . Now U dominates D for player 1, who eliminates it. In this case iterated dominance succeeds in reducing each player's set of available strategies to a singleton, so we can predict that the unique resulting profile of undominated strategies is the solution to the game. It is also easy to prove that such strategy profile is a Nash equilibrium (Bicchieri 1993). Note that assuming common knowledge of rationality (or at least some level of mutual knowledge of rationality) is crucial to obtaining the (U, L) solution. If there is some doubt that a player may not be fully rational - for example, if 1 thinks there is a 0.01 chance that R is chosen - then he does better by choosing D .

In **Figure 5** there are two Nash equilibria in pure strategies: (a, c) and (b, d) , and in the (a, c) equilibrium each player plays a weakly dominant strategy. Should we confidently predict that (a, c) will be the solution of the game? When a player has a strictly dominant strategy, rationality dictates choosing it, hence predicting behaviour is a simple matter. The case of weakly dominant strategies is not so straightforward. For one thing, it is not true that a weakly dominated strategy cannot be a best reply to any beliefs. In fact, for any weakly dominated strategy there is always some belief for which the strategy is a best reply. Putting the point differently, weak dominance means that there is at least one choice on the part of an opponent that makes one indifferent between the weakly dominated strategy and other strategies. In our example, were player 2 to believe that 1 plays b , her best reply would be d , and conversely, were player 1 to expect 2 to play d , his optimal strategy would be b .

		2	
		<i>c</i>	<i>d</i>
1	<i>a</i>	1, 1	0, 0
	<i>b</i>	0, 0	0, 0

Figure 5

Principles such as iterated strict dominance, rationalizability and Skyrms' (1990) deliberation dynamics are examples of how it is possible to restrict the set of predictions using rationality arguments alone. A very different approach is to start from the set of Nash equilibria, and ask whether some of them should be eliminated because they are in some sense 'unreasonable'. This is the approach taken by the *refinement* programme.

Normal form refinements. Consider again the game in Figure 5. How reasonable is the equilibrium (b, d) ? Under what circumstances would players agree to play it, and then stick to the agreement? If the problem is that players should not be too sure of the opponents' choices, one should try to model this uncertainty in the game. Selten's insight was to treat perfect rationality as a limit case (Selten 1965). His 'trembling hand' metaphor suggests that deciding and acting might be two separate processes, in that even if one decides to take a particular action, one may end up doing something else by mistake. More precisely, it is required that an equilibrium strategy is not only optimal against the opponents' strategies, but also some very small probability $\varepsilon > 0$ that the opponents make 'mistakes'. Such an equilibrium is *trembling-hand perfect*. Is the equilibrium (b, d) perfect? If so, b must be optimal against c being played with probability ε and d being played with probability $1 - \varepsilon$ for some small $\varepsilon > 0$. But the payoff to a is ε , while the payoff to b is 0. Hence for all $\varepsilon > 0$, a is a better strategy choice. The equilibrium (b, d) is not perfect, but (a, c) is. A prudent player will thus rule out (b, d) . In this simple game, checking the perfection of some action is easy, since only one 'mistake' is possible. With many actions, there are many more possible mistakes to take into account. Similarly, with many players we may need to worry about who is most likely to make a mistake. Note that the starting point of this approach is the set of Nash equilibria of the game. It is assumed that players will coordinate upon one of them, and thus the problem becomes that of ruling out all those equilibria that, for one reason or another, are not reasonable agreement points. Specifying why an equilibrium might be unacceptable is easier if one can consider what *would* happen were some of the players to 'deviate' from the agreed upon equilibrium. This is why much of the refinement literature refers to games in extensive form, where the dynamic structure of games, the order in which players move and the kind of information they have when making a choice are made explicit.

4 Games in extensive form

The *extensive form* of a game specifies the following information: a finite set of players $i = 1, \dots, n$, one of which might be nature (N); the order of moves; the players' choices at each move and what each player knows when he has to choose; the players' payoffs as a function of their moves; finally, moves by nature correspond to probability distributions over exogenous events. The order of play is represented by a game tree T , which is a finite set of partially ordered nodes $t \in T$ that satisfy a precedence relation denoted by ' $\prec$ '. The information a player has when he is choosing an action is represented using *information sets*, which partition the nodes of the tree. Since an information set can contain more than one node, the player who has to make a choice at an information set that contains, say, nodes t and t' will be uncertain as to which node he is at. If a game contains information sets that are not singletons, the game is one of *imperfect information*, in that one or more players will not know, at the

moment of making a choice, what the preceding player did. A *subgame* is a collection of branches of a game which start from the same node and which, together with the node, form a game tree by itself. In Figure 6, player 2's decision node as well as her moves form a subgame of the game. Finally, the games I shall consider here are all games of *perfect recall*, in that a player always remembers what she did and knew previously. A *strategy* for player i is a complete plan of action that specifies an action at every node at which it is i 's turn to move. Note that a strategy specifies actions even at nodes that will never be reached if that strategy is played. The game in Figure 6 is a finite game of perfect information where player 1 moves first. If he chooses D at the first node, the game ends and he nets a payoff of 1, whereas player 2 only gets 0. But choosing D at the first node is only *part* of a strategy of player 1; for example, it can be part of a strategy that says 'play D at the first node, and x at the last node'. It is important that a strategy specifies actions even at nodes that will not be reached: this allows players to say what *would* happen were those nodes to be reached by mistake or for some other reason. The game of Figure 6 has two Nash equilibria in pure strategies: (Dx, d) and (Dy, d) . This is easy to verify by looking at Figure 6', the normal form representation of the game. Is there a way to eliminate one of the two equilibria?

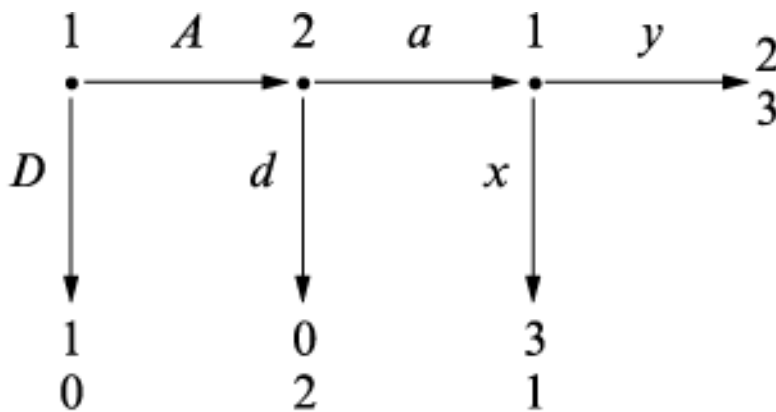


Figure 6

		2	
		<i>d</i>	<i>a</i>
1	<i>Dx</i>	1, 0	1, 0
	<i>Ax</i>	0, 2	3, 1
	<i>Dy</i>	1, 0	1, 0
	<i>Ay</i>	0, 2	2, 3

Figure 6'

Suppose player 1 were to play at his last node. Since he is rational, he would choose *x*, which guarantees him a payoff of 3. Knowing by assumption that 1 is rational, player 2 - if she were to reach her decision node - would play *d*, since by playing *a* she would net a lower payoff. Finally, since by assumption player 1 knows that 2 is rational and that she knows that 1 is rational, he will choose *D* at his first decision node. (Dy, d) should thus be ruled out, since it recommends an irrational move at the last node. In the normal form, iterated elimination of weakly dominated strategies allows both equilibria to survive. The reason is simple: if 1 plans to choose *D* and 2 plans to choose *d*, it does not matter what player 1 would do at the last node, since that node will never be reached. The procedure used in the extensive form to conclude that only (Dx, d) is a reasonable solution is referred to as *backwards induction* (Zermelo 1913). In finite games of perfect information in which there are no ties in payoffs, backwards induction always identifies a unique equilibrium. The use of backwards induction to solve our game presupposes that players analyse the game from the end up and that, once a subgame has been reached, the conclusion about what is rational to do in that subgame is independent of any consideration of how that subgame was reached. The strategic irrelevance of past moves - which is crucial to backwards induction - has been questioned by many philosophers and game theorists. If player 2 were to reach her decision node, would she keep thinking that player 1 is rational? Clearly her choice will depend on how she answers this question. Binmore (1987), Bicchieri (1989, 1993) and Reny (1992) were the first to point out that backwards induction may be inconsistent with common knowledge of rationality. The jury is still out and in recent years much foundational work has been devoted to model more precisely players' knowledge and reasoning in games.

Extensive form refinements. The goal of the refinement programme is that of eliminating implausible equilibria. In the normal form, Selten's trembling-hand perfection requires that players ask how a strategy will perform were another player to take an action that has zero probability in equilibrium. In the extensive form representation, we ask what would happen off-equilibrium, at points in the game tree that will never be reached if a given equilibrium is played. In both cases, the starting point is an equilibrium, which is checked for *stability* against possible deviations. By its nature, the Nash equilibrium concept does not restrict action choices off the equilibrium path, since those choices do not affect the payoff of the player who moves there. For example, the equilibrium (Dy, d) in Figure 6 lets player 1 make an irrational choice at the last node, since that choice is not going to affect his

payoff (which is determined by his choosing D at the beginning of the game). However, the strategy of a player at an off-equilibrium information set can affect what other players choose at their equilibrium information sets: in **Figure 6**, in order to choose D , player 1 will have to decide what would happen were he to play A instead. He thus has to think about player 2's choice at an off-equilibrium node. His conclusion about 2's choice will obviously affect his own move. *Subgame perfection* is a criterion proposed by Selten (1965) to rule out implausible equilibria in extensive games of perfect information. A subgame perfect equilibrium is a Nash equilibrium such that the strategies, when restricted to any subgame, remain a Nash equilibrium of the subgame. The equilibrium (Dy, d) is not subgame perfect. Note that in these games, the equilibrium obtained by backwards induction is also the unique subgame perfect equilibrium.

However, as **Figure 7** suggests, not all subgame perfect equilibria are sensible. In this game both (c, L) and (a, R) are trivially subgame perfect, since this game has no proper subgames. To obviate this problem, further refinements have been proposed. I will mention here only some of the most important. Kreps and Wilson (1982) introduced the concept of *sequential equilibrium*. Briefly, a sequential equilibrium is a strategy combination and a system of beliefs with the following properties: each player has a belief over the nodes at each of his information sets. Beginning at any information set, given a player's belief there and given the strategies of the other players, his own strategy for the rest of the game still maximizes his expected payoff. Beliefs must be obtained from the beliefs at earlier information sets and the equilibrium strategies by Bayes' rule. A *perfect equilibrium* (Selten 1975) is an equilibrium that is robust to small perturbations of players' equilibrium strategies. A *proper equilibrium* (Myerson 1978) is an equilibrium that is robust with respect to 'plausible' deviations, interpreted as deviations that do not involve costly mistakes. An alternative refinement that does not involve consideration of mistakes is the idea of *forward induction* (Kohlberg and Mertens 1986). The main idea is that off-equilibrium beliefs should be consistent with interpretations of deviations as rational moves, or *signals* that the deviating player intends to play something different.

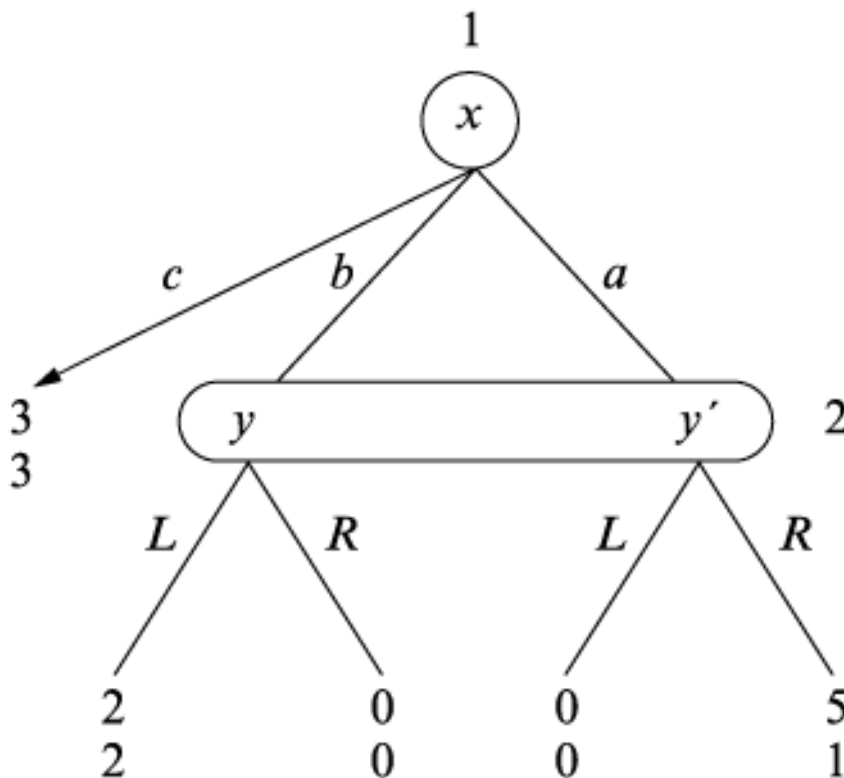


Figure 7

5 Selection by evolution

An alternative way of dealing with multiple equilibria is to suppose that the selection process is made by nature. In

an *evolutionary game* players are endowed with rules (strategies) and are randomly matched with other players in repeated encounters. A well known evolutionary model is Axelrod's (1984) prisoner's dilemma tournament in which strategies were matched against each other in turn. Survival of the fittest favours rules that are not exploitable by other rules and do well when played against themselves, because otherwise as the rule invaded the population and thus became more frequent, it would become weakened by lack of success. Maynard Smith (1982) has defined an *evolutionarily stable strategy* (ESS) as a strategy p^* which (i) is a best reply to itself and (ii) if there is any other strategy p that is a best reply to p^* , then p^* must be a better reply to p than p is to itself. As an example, consider a modified version of the game in Figure 3, one in which the payoffs for HD are (4,0), and those for DH are (0,4), and suppose that the two strategies are 'hawk' and 'dove'. A hawk always fights and escalates contests until it wins or is badly hurt. A dove never escalates; if it fights with another dove, they will settle the contest by sharing the resource. Suppose injury has a payoff in terms of loss of fitness equal to -6 . A victory nets 4. If hawk meets hawk or dove meets dove, each has a 50 per cent chance of victory. So a hawk meeting another hawk has an expected payoff of $1/2 \times 4 + 1/2 \times (-6) = -1$. If hawk meets dove, hawk gets 4 and dove gets 0; finally, two doves can expect a payoff of $1/2 \times 4 = 2$. The unique ESS turns out to be the mixed-strategy Nash equilibrium of the game: play hawk with probability $2/3$ and dove with probability $1/3$. If it were the case that 'hawkishness' and 'doveness' are genetic traits, we would expect the population to evolve so that two thirds of it is made up of hawks and one third of doves. Note that the evolutionary model gives a natural interpretation to mixed strategies as the proportions of certain strategies (or traits) in a population. This is an example of how evolution can select one equilibrium among many. In this particular game there are three equilibria - HD, DH, and $(1/3 D, 2/3 H)$ - but only one of them is evolutionarily stable.

6 Repeated games

Repeated games model ongoing interactions and are thus suitable models for phenomena like cooperation, trust, revenge, punishment and reputation formation. Take the prisoner's dilemma depicted in Figure 1, where C stands for 'cooperate' and D for 'defect'. Suppose this game is repeated N times. At each stage, players know the actions taken by all players at the previous stage, and the payoffs to each player are simply the sum of the payoffs he receives at each stage. Since the game is noncooperative, players cannot make enforceable agreements to cooperate. Moreover, a promise to cooperate or a threat to retaliate if the other does not cooperate, in order to be effective, must be credible. Both players know that in the last N th repetition both will defect, no matter what they did before. So they might as well defect in the $N - 1$ th repetition, too. Proceeding inductively, both players defect in each period, and indeed this is the unique subgame perfect equilibrium (see §4) of the repeated game. Consequently a promise to cooperate is not credible, nor is a threat to retaliate. They are not credible because they involve irrational behaviour, and rationality is common knowledge among the players.

When the game is infinitely repeated, the inductive argument that leads players to defect at each stage fails. In such games mutual cooperation is possible, indeed is one of many possible equilibria. This result is known as the *Folk theorem* - so called because it is part of the folk literature of game theory but its authorship is obscure (Aumann 1981). It states that any feasible expected payoff combination can be sustained in an equilibrium provided that each player has an expected payoff at least as large as she can guarantee for herself even if all other players pick strategies solely to punish her. In our example, any pair of feasible payoffs that gives each player more than 1 can be sustained in equilibrium. This result also holds for games in which players discount future payoffs, provided the discount rate is small. And it holds for games with an expected finite number of repetitions but a small probability that the game ends at each period.

As an example, consider indefinitely repeating the game in Figure 1, where players roll a fair dice after each stage and continue playing if and only if neither 5 nor 6 appears. There is a probability of $1/3$ that each stage is the last, and the probability that the game will continue until at least the N th stage is $(2/3)^{N-1}$. Since $(2/3)^N$ goes to zero as N goes to infinity, the probability that the game will go on forever is zero. Suppose player 1 announces that they will cooperate as long as player 2 is cooperative, but as soon as 2 defects, 1 will defect forever. The expected payoff of mutual cooperation (C) is $2 + 2(2/3) + \dots + 2(2/3)^N + \dots + 2(2/3)^{N+2} + \dots$. Can player 2 gain by deviating and, say, play D (defect) at the $N + 1$ th stage? In this case 2 will at most get $2 + 2(2/3) + \dots + 2(2/3)^{N-1} + 3(2/3)^N + 1(2/3)^{N+1} + 1(2/3)^{N+2} + \dots$. Since the expected payoff of C is better than the expected payoff of D , it is better for player 2 to cooperate. Mutual cooperation is a Nash equilibrium (see §3).

Cooperation can be achieved also in a finitely repeated game, provided players are not too sure about each other's rationality. Suppose that one player assigns a small probability that the other is a *tit-for-tat* player, that is, that he cooperates in the first period, and then mimic what the other player did in the preceding period. Kreps *et al.* (1982) showed that if the game is one of incomplete information (see §3), a small probability that one of the players is using the tit-for-tat strategy is enough to guarantee that both will cooperate until near the last period. Cooperation for several periods is a rational strategy for both players, provided that the prior probability of one of them playing tit-for-tat is common knowledge. This is an important result, as it explains at least some of the experimental evidence about cooperative behaviour in situations that have a known finite time-horizon. Furthermore, it shows that cooperative behaviour is possible, and compatible with individual rationality, even in the absence of external enforcing mechanisms.

7 Bargaining games

Consider a game in which two players must split three (indivisible) dollars between them. They must choose shares x_1, x_2 simultaneously. Their payoffs are defined as follows: if $x_1 + x_2 \leq 3$, each player gets the share he proposed; if $x_1 + x_2 > 3$, each gets nothing. There are two Nash equilibria in pure strategies: (1, 2) and (2, 1) in which one player demands one dollar and the other asks for two. One way to solve this indeterminacy is to allow the players to make a sequence of offers and counter-offers. For example, player 1 may be allowed to make the first offer; if it is accepted, player 2 gets the remainder and the game ends. If it is rejected, it is now player 2's turn to make an offer, and player 1 can accept or reject it. If he rejects, each player gets a payoff of zero. It is now possible to apply backwards induction, and predict that player 2 will offer one dollar to player 1 at the second stage, knowing that he will accept. Thus player 1 knows that at the first stage the only acceptable offer is two dollars to player 2. Further specification of the bargaining process thus determines the outcome of the bargaining problem as a subgame perfect equilibrium, in which player 1 gets one dollar and player 2 gets two dollars. The outcome, however, is very sensitive to the details of the process, in this case who moves first.

An alternative way to get a prediction for the outcome of a bargaining problem is to list a set of properties that an outcome is expected to have. Consider again our bargaining problem, but assume now the three dollars are divisible, and that players' marginal utility of money is positive and constant. As before, if the players do not reach an agreement, they get nothing, $(x_0, y_0) = (0, 0)$. Acceptable shares need not sum to \$3.

In the case of two-person bargaining, we can draw the set of possible money combinations (the outcome set) on a graph with x representing the share as well as the utility of player 1 on one axis, and y representing the share as well as the utility of player 2 on the other. In Figure 8, the origin has been placed at that utility combination that corresponds to total disagreement (x_0, y_0) .

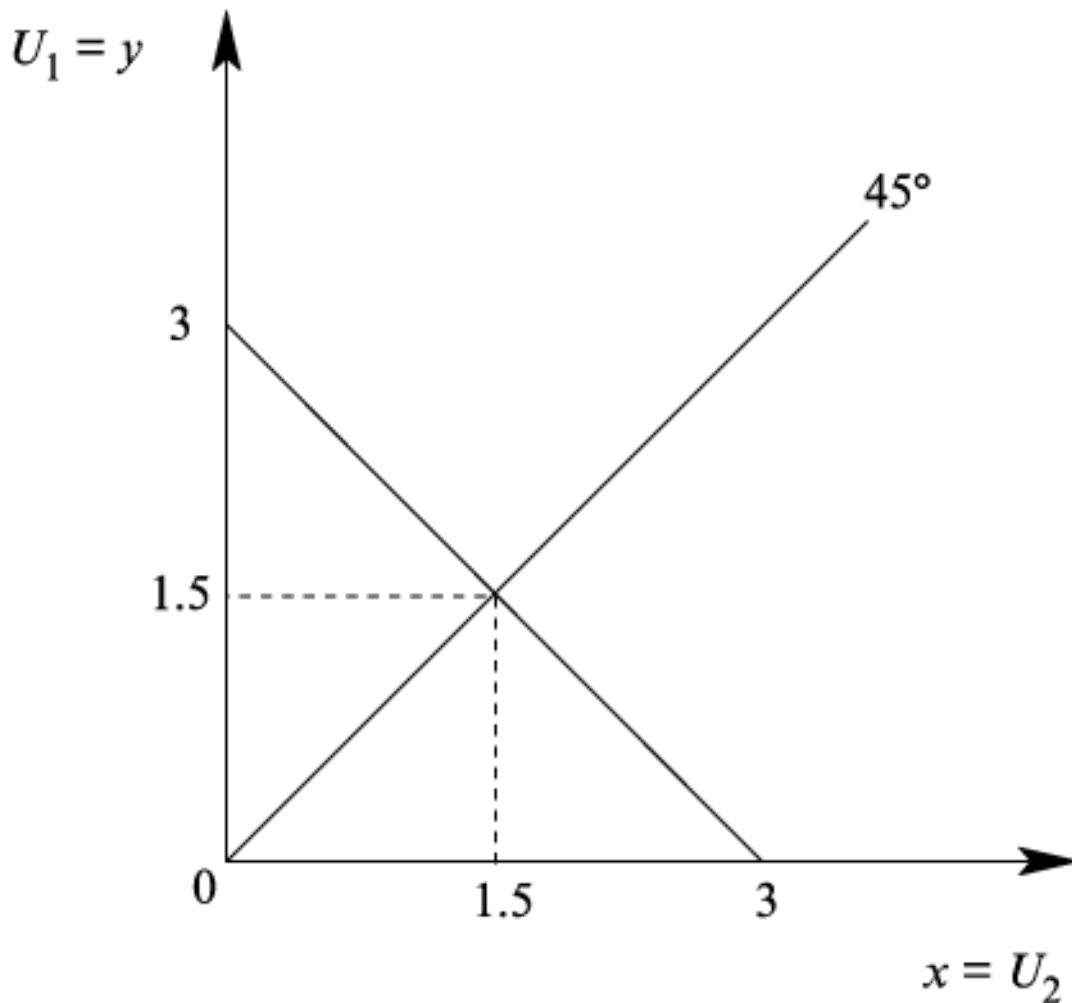


Figure 8

Nash (1950, 1953) suggested that it is reasonable to expect the outcome of a bargaining process to satisfy the following conditions:

- (1) *Pareto optimality*: whatever agreement is reached, it will not be possible to improve it without damaging one of the players.
- (2) *Independence of irrelevant alternatives*: if we drop some possible utility combinations from the outcome set, and if the actual outcome of the original game is also a possible outcome in the restricted game, then it will be the final outcome for the restricted game as well.
- (3) *Symmetry*: if the set of possible outcomes is symmetric (that is, if for every point $x = a$, $y = b$ in the outcome set, the point $x = b$, $y = a$ is also in the set), then the outcome of bargaining will give $x = y$.
- (4) *Invariance*: the solution is independent of the units in which we measure utility.

Nash proved that there is a unique point that satisfies the four conditions, namely the utility combination (x^*, y^*) such that $(x^* - x_0)(y^* - y_0) \geq (x - x_0)(y - y_0)$ for all (x, y) in the outcome set.

The bargaining set of our example is symmetric, hence the outcome must lie on the 45°-line, and Pareto optimality requires it to be on the Pareto frontier. This identifies $(x^*, y^*) = (1.5, 1.5)$ as the unique outcome. Clearly our prediction of the outcome of a bargaining problem will depend upon our degree of confidence in the conditions that a solution should satisfy. There is, however, a link between the latter approach and the strategic one: Nash himself suggested that any set of conditions that identifies a particular outcome should be supplemented by a bargaining process that results in that outcome.

8 Cooperative games

In a noncooperative game, there is no way for the players to enter into binding agreements, or to commit themselves to playing a particular strategy. Agreements are of course possible in such games, but to be honoured they have to be self-enforcing, that is, they must be Nash equilibria. In a cooperative game on the other hand, agreements between the players are enforceable. The Nash bargaining solution introduced an approach to games that overlooks moves and strategies in favour of considering the set of all possible outcomes of an interaction and selecting a particular outcome based on some prior conditions. This approach makes sense if we allow for the possibility of enforcing contracts, since in this case every outcome that can be jointly achieved by the players is a feasible agreement.

While the Nash solution can be generalized to n players (Harsanyi 1977), the possibility of coalition formations among the players raises new problems. Games in which coalitions can arise are described in *characteristic function* form. Such games are completely described by the players, the coalition structure, and the specification of what each coalition can guarantee itself. An example of a game in characteristic function form is the following: there are three players who must divide a dollar between themselves with the convention that a majority of two rules. Since a majority can decide on how to split the dollar, a single player cannot guarantee any amount. Two players can together guarantee a total of one dollar, and the coalition consisting of all the players (the grand coalition) can guarantee itself one dollar. We define an *imputation* as a vector $x = \{x_1, \dots, x_n\}$ such that $\sum_{i \in N} x_i = V(N)$, the total amount the grand coalition can guarantee itself if there is a total of N players; in the game above this sum is one dollar. We also require that each x_i must be at least as great as what player i can guarantee himself. The imputation thus refers to the amount each player receives in a given outcome. It does not yet indicate the solution of the game. A solution for such games is a proposed distribution of the joint payoff to the players that is accepted. There are many ways to characterize payoff vectors as acceptable. One approach is to provide criteria that an acceptable allocation should satisfy. Two such criteria are Pareto optimality and symmetry. The *Shapley value* (Shapley 1953) is the unique payoff allocation that satisfies the above criteria. It is such that the value of each player i in a game turns out to be the average of all marginal contributions that player i can make to all coalitions.

A natural requirement for any payoff vector to be mutually agreeable seems to be that no subgroup could do better by breaking the agreement. The core concept captures this idea formally. The *core* consists of all imputations satisfying $\sum_{i \in S} x_i = V(S)$, the amount coalition S can guarantee, for all possible S . In the dollar division game the core is empty, indicating that there is no allocation that is stable against deviations by subgroups.

See also: [Behaviourism in the social sciences](#); [Economics, philosophy of](#); [Probability, interpretations of](#); [Rational beliefs](#); [Rational choice theory](#); [Semantics, game-theoretic](#); [Social action](#)

CRISTINA BICCHIERI

References and further reading

- Allais, M. (1953) 'Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école américaine', *Econometrica* 21: 503-46. (Shows how the von Neumann-Morgenstern Independence axiom is sometimes violated.)
- Arnould, A. (1662) *La Logique, ou l'art de penser*, Paris; 2nd edn, Amsterdam: Wolfgang, 1685. (A very early theory of decision-making under uncertainty. See specifically chapter XVI, part III.)
- Aumann, R. (1974) 'Subjectivity and Correlation in Randomized Strategies', *Journal of Mathematical Economics* 1: 67-96. (Introduces the concept of *correlated equilibrium*.)
- Aumann, R. (1976) 'Agreeing to Disagree', *Annals of Statistics* 4: 1,236-9. (Defines *common knowledge* in games.)
- Aumann, R. (1981) 'Survey of repeated games', in R. Aumann (ed.) *Essays in Game Theory and Mathematical Economics in Honor of Oscar Morgenstern*, Mannheim: Bibliographisches Institut. (A good introduction to repeated games.)
- Axelrod, R. (1984) *The Evolution of Cooperation*, New York: Basic Books. (Shows how a cooperative strategy can do well in computer tournaments, and how cooperation can evolve in a mixed population.)
- Bernheim, B.D. (1984) 'Rationalizable Strategic Behavior', *Econometrica* 52: 1,007-28. (With Pearce's (1984) article, the classic definition of *rationalizability*.)
- Bernoulli, D. (1738) 'Specimen theoriae novae de mensura sortis', *Commentarii Academiae Scientiarum*

- Imperialis Petropolitanae* V: 75-92; trans. L. Sommer, 'Exposition of a New Theory on the Measurement of Wealth', *Econometrica* 22: 23-36. (Provides the oldest definition of expected utility and applies it to solving the St Petersburg paradox.)
- Bicchieri, C.** (1989) 'Self-Refuting Theories of Strategic Interaction: A Paradox of Common Knowledge', *Erkenntnis* 30: 69-85. (Shows how common knowledge of rationality is inconsistent with backwards induction.)
- Bicchieri, C.** (1993) *Rationality and Coordination*, Cambridge: Cambridge University Press. (A comprehensive introduction to the philosophical foundations of game theory.)
- Binmore, K.** (1987) 'Modeling Rational Players: I', *Economics and Philosophy* 3: 179-214. (A criticism of traditional justifications of backwards induction and a reflection on the meaning of rationality in strategic contexts.)
- Butler, J.** (1736) *The Analogy of Religion, Natural and Revealed, to the Constitution and Course of Nature*, London. (Classic text referred to in §1.)
- Ellsberg, D.** (1961) 'Risk, Ambiguity, and the Savage Axioms', *Quarterly Journal of Economics* 75: 643-69. (Documents a famous violation of the von Neumann and Morgenstern Independence axiom.)
- Fishburn, P.C.** (1982) *The Foundations of Expected Utility*, Dordrecht: Reidel. (Comprehensive introduction to decision theory.)
- Harsanyi, J.** (1967) 'Games with Incomplete Information Played by Bayesian Players', *Management Science* 14: 159-82, 320-34, 486-502. (Reduces games of incomplete information to games of complete information.)
- Harsanyi, J.** (1973) 'Games with Randomly Disturbed Payoffs: A New Rationale for Mixed Strategy Equilibrium Points', *International Journal of Game Theory* 2: 1-23. (Introduces the idea of *purification* of mixed strategies.)
- Harsanyi, J.** (1977) *Rational Behaviour and Bargaining Equilibrium*, Cambridge: Cambridge University Press. (A comprehensive treatise on bargaining games.)
- Kahneman, D. and Tversky, A.** (1979) 'Prospect Theory: An Analysis of Decision Under Risk', *Econometrica* 47: 263-91. (A classical criticism of decision theory based on experimental evidence.)
- Kohlberg, E. and Mertens, J.-F.** (1986) 'On the strategic stability of equilibria', *Econometrica* 54: 1,003-37. (Introduces the idea of *forward induction* as a refinement of Nash equilibrium.)
- Kreps, D. and Wilson, R.** (1982) 'Sequential Equilibria', *Econometrica* 50: 863-94. (Defines the concept of *sequential equilibrium* as a refinement of Nash equilibrium.)
- Kreps, D., Milgrom, P., Roberts, J. and Wilson, R.** (1982) 'Rational Cooperation in the Finitely Repeated Prisoner's Dilemma', *Journal of Economic Theory* 27: 245-52. (Shows how incomplete information allows players to cooperate in a finitely repeated prisoner's dilemma.)
- Lewis, D.** (1969) *Convention: A Philosophical Study*, Cambridge, MA: Harvard University Press. (A study of conventions as coordination equilibria.)
- Machina, M.** (1982) 'Expected Utility Analysis Without the Independence Axiom', *Econometrica* 50: 277-323. (An alternative axiomatization of expected utility.)
- Maynard Smith, J.** (1982) *Evolution and the Theory of Games*, Cambridge: Cambridge University Press. (A classic application of game theory to biology.)
- Myerson, R.** (1978) 'Refinements of the Nash Equilibrium Concept', *International Journal of Game Theory* 7: 73-80. (Introduces the concept of *proper equilibrium*.)
- Nash, J.** (1950) 'The Bargaining Problem', *Econometrica* 18: 155-62. (Introduces the *Nash bargaining solution*.)
- Nash, J.** (1951) 'Noncooperative Games', *Annals of Mathematics* 54: 289-95. (Defines and proves the existence of *Nash equilibrium*.)
- Nash, J.** (1953) 'Two-person Cooperative Games', *Econometrica* 21: 128-40. (Proves there exists a solution for such games that satisfies certain desirable properties.)
- Neumann, J. von** (1928) 'Zur theorie der gesellschaftspiele', *Mathematische Annalen* 100: 295-320. (Contains the proof of the *minimax theorem*.)
- Neumann, J. von and Morgenstern, O.** (1944) *Theory of Games and Economic Behavior*, Princeton, NJ: Princeton University Press. (The classic, if outdated, introduction to games theory.)
- Osborne, M. and Rubinstein, A.** (1994) *A Course in Games Theory*, Cambridge, MA: MIT Press. (An excellent, comprehensive textbook.)
- Pearce, D.G.** (1984) 'Rationalizable Strategic Behavior and the Problem of Perfection', *Econometrica* 52: 1,029-50. (Introduces the concept of *rationalizability* independently of Bernheim.)
- Reny, P.** (1992) 'Rationality in Extensive Form Games', *Journal of Economic Perspectives* 6: 92-100. (Shows how common knowledge assumptions may be inconsistent with backwards induction.)
- Savage, L.** (1954) *The Foundations of Statistics*, New York: Dover. (Seminal work on decisions under

uncertainty.)

Schelling, T. (1960) *The Strategy of Conflict*, Cambridge, MA: Harvard University Press. (Contains a simple but comprehensive analysis of coordination games.)

Selten, R. (1965) 'Spieltheoretische Behandlung eines Oligopolmodells mit Nachfrageträgheit', *Zeitschrift für die gesamte Staatswissenschaft* 121: 301-24. (Introduces the idea of *trembling hand perfection* for games in normal form.)

Selten, R. (1975) Re-examination of the Perfectness Concept for Equilibrium Points in Extensive Games', *International Journal of Game Theory* 4: 22-55. (Introduces the concept of *perfect equilibrium*.)

Shapley, L.S. (1953) 'A Value for n-person Games', in H.W. Kuhn and A.W. Tucker (eds) *Contributions to the Theory of Games*, Princeton, NJ: Princeton University Press, vol. II, 307-17. (Introduces the now classical solution concept for cooperative games.)

Skyrms, B. (1990) *The Dynamics of Rational Deliberation*, Cambridge, MA: Harvard University Press. (See §3. Deliberation dynamics, an example of how it is possible to restrict the set of predictions using rationality arguments alone.)

Zermelo, E. (1913) 'Über eine Anwendung der Mengenlehre auf die Theorie des Schachspiels', in *Proceedings of the Fifth International Congress of Mathematicians*, vol. II, Cambridge: Cambridge University Press. (The original statement of the *backwards induction* algorithm.)

Definition

A definition is a statement, declaration or proposal establishing the meaning of an expression. In virtue of the definition, the expression being defined (the 'definiendum') is to acquire the same meaning as the expression in terms of which it is defined (the 'definiens'). For example, 'Man is a rational animal' determines the meaning of the term 'man' by making it synonymous with 'rational animal'. Classical theory maintains that a good definition captures the 'real nature' of what is defined: 'A "definition" is a phrase signifying a thing's essence' (Aristotle). Historically, philosophers have come to distinguish these 'real' definitions from 'nominal' definitions that specify the meaning of a linguistic expression rather than signify the essential nature of an object, 'making another understand by Words, what Idea, the term defined stands for' (Locke).

A further distinction can be drawn between contextual or implicit definitions, on the one hand, and explicit definitions, on the other. Often a definition fixes meaning directly and explicitly: for example, the definition of a proper name might well take the form of an explicit identity statement ('Pegasus = the winged horse') and a definition of a predicate is usually given (or can be re-cast) in the form of an equivalence ('For every x: x is a man if and only if x is a rational animal'). But sometimes the meaning of a term is specified in context, by way of the meaning of larger expressions in which the term occurs. A paradigmatic example of this is Bertrand Russell's analysis of the meaning of the definite article.

1 The role of definitions

A definition is a statement, declaration or proposal (the 'definiens') establishing the meaning of an expression (the 'definiendum'). Explicit definitions can be viewed as providing useful abbreviations for complex expressions. It follows that, strictly speaking, they are theoretically dispensable, in so far as we can always replace the definiendum by the definiens in any context. This was already noted by Whitehead and Russell in *Principia Mathematica*, according to whom definitions are 'mere typographical conveniences' ([1910] 1925: 11). However, they also point out that from a more pragmatic point of view, the use of a definition (whether explicit or not) conveys additional information besides the literal meaning of the sentence in which the definiendum occurs.

This can happen in two ways. First, a definition calls attention to the fact that the (complex) concept expressed by the definiens plays an important role in the discourse in which the definiendum is used: for instance, agreeing to mean by 'prime' a number that can be divided without remainder only by 1 and itself calls attention to the importance of this notion for number theory. 'Hence the collection of definitions embodies our choice of subjects and our judgment as to what is most important' ([1910] 1925: 11-12). Second, a definition may sometimes supply an analysis of an already existing concept that had not previously been made precise. In this case the newly introduced concept shares most or all of the traits of its informal counterpart, although realizing those traits in a more specific and precise manner. It should be observed that in general many competing, equally precise and sometimes formally equivalent analyses of the same informal idea are possible, and the choice among them is then inspired by considerations such as simplicity, elegance and adaptability to the particular context of application.

A related view of definitions was put forward by Carnap (1947), who characterizes certain definitions as 'rational reconstructions' of less precise concepts derived from everyday life or earlier stages of scientific development. Such a definition is called an 'explication', and the newly defined term the 'explicatum'; the earlier, less precise concept is referred to as the 'explicandum'. In giving an explication, one 'sharpens' a vague, pre-existing concept, making it precise; or even replaces it with a new, more exact one. In general it is not required that the explicandum have the same meaning as the explicatum. All that is required is some sort of correspondence between the two concepts, in such a way that the latter can be used instead of the former. This task of rational reconstruction is 'one of the most important...of logical analysis and logical construction' ([1947] 1956: 7-8).

2 Formal conditions on definability

From a purely formal point of view, a definition given in the context of some theory behaves just like any other axiom of the theory, in that it can be used to infer propositions (theorems). However, as already noted by Whitehead and Russell, a definition should not convey any new information: the theory T' obtained by supplementing some theory T by a definition should, in some sense, be equivalent to T .

It is possible to make this requirement precise by laying down two criteria that a definition should satisfy: eliminability and non-creativity. It appears that these criteria were first formulated by the Polish logician **S. Leśniewski (1931)**, but the presentation given here follows **P. Suppes (1957)**.

In order to explain the two criteria, we will suppose that we are working within a formalized background theory T , that is, a set of axioms expressed in some formal language. In such a language, as is customary, free variables are to be construed as implicitly universally quantified; moreover, T is assumed to be consistent. A definition then is just a sentence ϕ that involves some new symbol not already in T . In particular, ϕ is allowed to occur in inferences from the axioms of T . The two criteria can be formulated as follows. A sentence ϕ introducing a new symbol satisfies the criterion of eliminability if and only if whenever η is a formula in which the new symbol occurs, then there is a formula θ in which the new symbol does not occur and such that $(\eta \leftrightarrow \theta)$ can be inferred from T augmented with ϕ . Similarly, we say that ϕ satisfies the criterion of non-creativity if and only if whenever θ is a sentence not involving the new symbol and such that $\phi \rightarrow \theta$ can be inferred from T , then θ can already be inferred from T ; in other words, the introduction of the definition does not make possible the derivation of any new theorems that do not involve the new symbol. (One then says that $T + \phi$ is a ‘conservative extension’ of T .) Notice that the second criterion implies, in particular, that the theory supplemented by the definition be free of contradictions: inconsistency is a strong form of creativity.

Although the two criteria apply to explicit and implicit definitions alike, in the case of explicit definitions it is possible to lay down certain constraints on the syntactic form of the sentence ϕ ensuring that the two criteria are automatically satisfied. For simplicity, we will review these constraints only for the case in which the definiendum is a relation symbol or an individual constant (name), but a similar constraint can be formulated for the definition of function symbols.

A formula ϕ is an acceptable definition of an n -place relation symbol R if and only if ϕ has the form $R(x_1, \dots, x_n) \leftrightarrow \theta$, where $x_1, \dots, x_n$ are all distinct variables and θ is a formula all of whose free variables are among $x_1, \dots, x_n$ and in which R does not occur. The requirement that the variables $x_1, \dots, x_n$ all be distinct is necessary to ensure that the eliminability criterion is satisfied, as is the further requirement that the definiendum R not be allowed to occur in the definiens; jointly, the two requirements are also sufficient for the satisfaction of the eliminability criterion.

The requirement that the definiens not contain any variables that do not already occur in the definiendum is meant to ensure non-creativity. To see that the constraint expresses a necessary condition, consider, for instance, the definition ‘ $G(x) \leftrightarrow x > y$ ’. Since free variables are implicitly universally quantified, the definition is equivalent to the conjunction of ‘ $G(x) \rightarrow \forall y(x > y)$ ’ and ‘ $\exists y(x > y) \rightarrow G(x)$ ’; by transitivity of implication we obtain that if $\exists y(x > y)$ then $\forall y(x > y)$, which, when interpreted as a statement of arithmetic, is false. The definition is therefore creative. (The constraint on variables is also sufficient for non-creativity if and only if the definiendum is not allowed to occur in the definiens.)

The constraint on the explicit definition of individual constants is slightly more complex, due to the existential import of such constants in ordinary logic. Rather than considering definitions in the form of an explicit identity ‘ $c = \text{the so-and-so}$ ’, we will give a constraint using the more general form of an equivalence. A sentence ϕ is an acceptable definition of an individual constant c if and only if ϕ has the form $c = x \leftrightarrow \theta$, where θ is a formula not containing c with x as its only free variable, and such that $\exists! x \theta$ (that is, there is one and only one x such that θ) is derivable from the theory.

3 Contextual definitions

Sometimes it is neither possible nor convenient to give a definition in the explicit form of an equivalence or identity. Indeed, given that formally definitions are just further axioms, any sentence containing a new symbol could be taken to be a candidate for a definition of that symbol. Such definitions are called ‘contextual’, in that the meaning of the new symbol is specified (at least partially) by giving the meaning of some larger expression in which it occurs; the symbol itself, when occurring in isolation, need not be endowed with any meaning at all. Perhaps the first and most conspicuous example of contextual definition is given by Russell (see, for example, Whitehead and Russell [1910] 1925: 66; but see also Frege 1884, 1903), when specifying the meaning of the definite article by giving explicit truth-conditions for sentences in which it occurs. As is well known, Russell

proposes to analyse sentences of the form ‘The ϕ is ψ ’ as the conjunction of the three following sentences in which the definite article no longer occurs: (1) There is at least one ϕ ; (2) There is at most one ϕ ; and (3) Whatever is a ϕ is a ψ . Other examples are the contextual definition of the membership symbol ‘ $\in$ ’ given by the axioms of set theory, and the usual recursive definition of ‘+’ in the context of arithmetic (the latter is given by the conjunction of the two equations $x + 0 = x$ and $x + y' = (x + y)'$, where x' denotes the successor of x).

Of course, in general, one would not expect the criterion of eliminability to be satisfied, for a contextual definition need only tell us how to find an equivalent not involving the new symbol for a restricted class of sentences, namely those in which the definiendum occurs in the context given by the definition. For instance, the usual recursive definition of ‘+’ does not tell us how to eliminate this symbol in the context $\forall x \forall y (x + y = y + x)$. But the criterion of non-creativity still applies, along with the weaker requirement that the theory augmented by the definition still be consistent.

As a related question one might consider whether, within the framework of a particular theory, a contextual definition can be replaced by an explicit one. A useful tool in answering this question is a method invented by A. Padoa (1902) and further developed by Tarski (1956: 296-320). The method relies on the fact that if a symbol α is explicitly definable within a given theory T , then any two interpretations on which all sentences of T are true must agree on the interpretation of α . Therefore, in order to show that α is not explicitly definable in T it suffices to exhibit two interpretations of T that assign different values to α .

Padoa’s method also has a converse of sorts, which is a deep result about the expressive power of first-order logic. Let us say that a symbol α is ‘implicitly definable’ in T if any two interpretations of T agree on the value they assign to α . (So implicit definitions are a particular case of contextual definitions.) The proposition that any symbol implicitly definable in T is also explicitly definable is known as Beth’s definability theorem (see Beth’s theorem and Craig’s theorem). This is a powerful theorem, and one of the results of model theory with the most far-reaching philosophical implications.

4 ‘Circular’ definitions

We have seen that one of the constraints on the syntactic form of definitions meant to ensure the satisfaction of the criterion of eliminability is that the definiendum itself not occur in the definiens. That is, the requirement rules out ‘circular’ definitions. This is in keeping with the idea that a definition, among other things, should also provide a way to determine the extension of the defined term, that is, the set of entities that fall under it. Indeed, a way of determining whether an object x falls in the extension of, say, a predicate P is to ascertain whether x satisfies the definiens of P : if the definiens itself is allowed to have occurrences of P , then it might be impossible to determine whether anything satisfies it without knowing the extension of P . So circularly defined terms would lack a definite criterion of applicability.

However, there are circumstances in which it appears that no non-circular definition is available for a given concept. Consider, for instance, the concept of truth, construed as a predicate of sentences of a given language. If the language in question itself contains the truth-predicate then we can have circularity as follows. Tarski (1944) suggested that for any sentence ϕ , the following equivalence (known as a ‘T-biconditional’) might be regarded as a *partial* definition of truth: t is true if and only if ϕ , where t is a term denoting ϕ (for instance, t could be obtained by enclosing ϕ in quotes). If ϕ itself contains occurrences of ‘is true’, the T-biconditional for ϕ exhibits a form of circularity. As is well known, this circularity can lead to paradoxes, as in the case of the so-called liar paradox arising from the consideration of the T-biconditional for the sentence λ : λ is not true. As another example, consider the kind of (harmless) circularity that arises when we are trying to specify a function f by means of the two equations $f(x, 0) = x$ and $f(x, y') = (f(x, y))'$, where as before x' is the successor of x (this is harmless because, as is well known, the two equations above identify a unique function over the natural numbers, namely addition).

A general theory of possibly circular definitions was put forward by A. Gupta and N. Belnap (1993). They regard a definition not as supplying a way to fix the extension of a term, but as providing a ‘rule of revision’, that is, a way of improving upon hypothetical extensions. Suppose we have a possibly circular definition of a 1-place predicate P , which has the form $P(x) \leftrightarrow \phi(P)$, where the definiendum P occurs in the definiens. This can be viewed as an operation defined over the space of the possible extensions of P (that is, over the set of all subsets of the domain):

given a possible extension E as input the rule of revision returns the value of $\phi(P)$ when P is interpreted as E . In turn, a revision rule can be used to give rise to a revision process, that is, an infinite sequence of possible extensions, defined as follows. We begin with an arbitrary extension E as a ‘bootstrapper’ or initial guess; we then let E_1 be the result of applying the revision rule to E , E_2 the result of applying the revision rule to E_1 , and so on. In this way we obtain an infinite sequence. Say that an extension E' ‘coheres’ with such a sequence if it contains all items that are eventually in every member of the sequence, that is, all items x such that for some E_n , x is in E_i for every $i \geq n$. We can then extend the sequence into the transfinite by letting E^∞ come after each member of the sequence, where E^∞ is an arbitrary extension that coheres with the sequence, and start applying the revision rule again.

So each revision rule specifies a class of revision processes, each one of which is characterized by a different ‘bootstrapper’ and different ‘guesses’ at limit stages. Should a revision process culminate in a fixed point, that is, an extension that is returned unchanged by the revision rule, we have found a ‘good’ candidate for the extension of P . But it should be noted that this is a by-product of the definition: the meaning of the definition is not given by the extension (which might not exist), but by the pattern of variation exhibited by the various revision processes to which it gives rise.

See also: Frege, G. §§6, 8, 9; Logical and mathematical terms, glossary of; Paradoxes of set and property §3

G. ALDO ANTONELLI

References and further reading

- Belnap, N.D.** (1993) ‘On Rigorous Definitions’, *Philosophical Studies* 72: 115-46. (Expanded treatment of the issues covered here; rigorous yet accessible both from a philosophical and technical point of view; emphasizes definition as an act of conceptual clarification.)
- Carnap, R.** (1947) *Meaning and Necessity: A Study in Semantics and Modal Logic*, Chicago, IL: University of Chicago Press; 2nd, enlarged edn, 1956. (Presents Carnap’s notion of explication.)
- Frege, G.** (1884) *Die Grundlagen der Arithmetik: eine logisch-mathematische Untersuchung über den Begriff der Zahl*, Breslau: Koebner; trans. J.L. Austin, *The Foundations of Arithmetic: A Logico-Mathematical Enquiry into the Concept of Number*, Oxford: Blackwell, 2nd edn, 1980. (Numbers are defined contextually - as extensions of particular concepts - by giving their identity conditions.)
- Frege, G.** (1903) *Grundgesetze der Arithmetik: begriffsschriftlich abgeleitet*, Jena: Pohle, vol. 2; fragments, including ‘Frege on Definitions’, in *Translations from the Philosophical Writings of Gottlob Frege*, trans. and ed. P.T. Geach and M. Black, Oxford: Blackwell, 3rd edn, 1980. (Gives the ‘sharp boundary’ condition, emphasizing that definitions should be complete and not leave any cases undecided. Also, the inconsistent axiom (V) can be viewed as a (creative!) contextual definition of the extension of a concept.)
- Gupta, A. and Belnap, N.D.** (1993) *The Revision Theory of Truth*, Cambridge, MA: MIT Press. (Includes the first full presentation of revision theory as a general theory of definition.)
- Leśniewski, S.** (1931) ‘Über Definitionen in der sogenannten Theorie der Deduktion’, *Comptes Rendus des Séances de la Société des Sciences et des Lettres de Varsovie* (Classe 3) 24; trans. in *Collected Works*, ed. S.J. Surma, J.T. Szrednicki and D.I. Barnett, with an annotated bibliography by V.F. Rickey, Boston, MA: Reidel, 2 vols, 1981, 629-48. (The original source for the theory of definition in first-order logic.)
- Padoa, A.** (1902) ‘Un nouveau système irréductible de postulats pour l’algèbre’, in *Deuxième Congrès International des Mathématiciens*, Paris, 249-56. (First appearance of a necessary condition for implicit definability, later developed by Tarski.)
- Robinson, R.** (1950) *Definition*, Oxford: Oxford University Press. (A good source for the nontechnical issues in the theory of definition. Includes some history.)
- Suppes, P.** (1957) *Introduction to Logic*, Princeton, NJ: Van Nostrand. (Possibly the best treatment of definitions in first-order logic from an elementary point of view.)
- Tarski, A.** (1944) ‘The Semantic Conception of Truth’, *Journal of Philosophy and Phenomenological Research* 4: 341-76; repr. in L. Linsky (ed.) *Semantics and the Philosophy of Language: A Collection of Readings*, Champaign, IL: University of Illinois Press, 1969. (Tarski’s more philosophical treatment of the concept of truth, including the partial definition of truth as given by the T-biconditionals.)
- Tarski, A.** (1956) *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*, trans. and ed. J.H. Woodger, Oxford: Clarendon Press; repr. and ed. J. Corcoran, Indianapolis, IN: Hackett Publishing Company, 2nd edn, 1983. (Tarski’s development of Padoa’s method. See also his remarks on formal and material correctness in ‘The Concept of Truth in Formalized Languages’.)

Whitehead, A.N. and Russell, B.A.W. (1910) *Principia Mathematica*, vol. 1, Cambridge: Cambridge University Press, 2nd edn, 1925.(The source for contextual definitions - sums up Russell's work published elsewhere.)

Demonstratives and indexicals

Demonstratives and indexicals are words and phrases whose interpretations are dependent on features of the context in which they are used. For example, the reference of 'I' depends on conditions associated with its use: as you use it, it refers to you; as I use it, it refers to me. In contrast, what 'the inventor of bifocals' refers to does not depend on when or where or by whom it is used. Among indexicals are the words 'here', 'now', 'today', demonstrative pronouns such as 'this', reflexive, possessive and personal pronouns; and compound phrases employing indexicals, such as 'my mother'. C.S. Peirce introduced the term 'indexical' to suggest the idea of pointing (as in 'index finger').

The phenomenon of indexicality figures prominently in recent debates in philosophy. This is because indexicals allow us to express beliefs about our subjective 'place' in the world, beliefs which are the immediate antecedents of action; and some argue that such beliefs are irreducibly indexical. For example, my belief that I am about to be attacked by a bear is distinct from my belief that HD is about to be attacked by a bear, since my having the former belief explains why I act as I do (I flee), whereas my having the latter belief explains nothing unless the explanation continues 'and I believe that I am HD'. It seems impossible to describe the beliefs that prompt my action without the help of 'I'. Similarly, some have argued that indexical-free accounts of the self or of consciousness are necessarily incomplete, so that a purely objective physicalism is impossible. In a different vein, some (such as Putnam 1975) have argued that our terms for natural substances, kinds and phenomena ('gold', 'water', 'light') are indexical in a way that entails that certain substantive scientific claims - for example, that water is H_2O - are, if true, necessarily true. Thus, reflection on indexicality has yielded some surprising (and controversial) philosophical conclusions.

1 Semantic data

An indexical is context-dependent: its interpretation varies with changes in aspects of the context of use. Yet almost any expression exhibits variation in interpretation with respect to some parameter or other. For example, a contingent sentence is one which is possibly true and possibly false, so we may think of it as varying in truth-value with changes in possible worlds. But mere contingency is not a species of indexicality. Similarly, tensed verbs induce variation in interpretation with respect to the passage of time, but tense is not usually viewed by linguists and logicians as a kind of indexicality (although there is some confusion about the matter). The difference between true context dependence and mere 'parameter dependence' can be seen as follows. Consider the sentence

(1) Jones will remember what Smith said just now.

The truth or falsity of this sentence at a moment in time t depends on the truth or falsity of the present tense sentence

(2) Jones remembers what Smith said just now

at moments t' future to t . But the referent of the occurrence of 'now' in (2) must refer back to t and not to any of the future moments t' . This means that the semantic rule for 'now' cannot be simply the rule which states that at any time t , 'now' refers to t . Instead, the rule must say that a use of 'now' at a time t will refer at time t and at any other time to t . Thus, the rule for 'now' must be 'double-indexed'. In general, the semantic rules for indexicals must involve the pairing of two sorts of parameter: contextual and circumstantial. The contextual parameters (speaker, time of utterance, place of utterance, world of utterance, addressee and so on) fix the referent of an indexical, so that its interpretation is unaffected by any circumstantial parameter (possible world, moment of time) required by the presence of tense or modal elements. Hereafter, the phrase 'context-dependent' will refer to the dependence of indexicals on both context and circumstance.

Indexicals are 'referential' in that they invariably have wide scope when embedded in tensed, modal or propositional attitude constructions. For example, consider the two statements 'The mayor might not have been a Democrat' and 'That person over there, the Mayor, might not have been a Democrat'. The former statement could be interpreted to mean merely that the office of mayor might have been occupied by a non-Democrat; but the latter statement can only be interpreted to mean that a certain person (who happens to be the mayor) might not have been

a Democrat. This property of indexicals is an effect of their context dependence. The referent of the indexical phrase ‘that person’ is fixed by the context and not by circumstances required to interpret the modal element ‘might have been’.

Though the interpretation of an indexical will vary with the context, indexicals are not in general ambiguous. Your use of ‘I’ and mine are linked by a common meaning. The language supplies a rule - the same rule for both of us - for the use of ‘I’. The rule is roughly this: ‘As *x* uses “I” in such and such a context, “I” refers to *x* in all circumstances’. But this sort of meaning is not like the Fregean sense of a definite description or proper name; it is not a concept of some fixed object and it cannot be combined with the sense of a (non-indexical) predicate to yield a complete proposition (see Frege, G. §§2-3). Some theorists believe that this fact requires the introduction into semantics of a whole new level of meaning distinct from both sense and reference. Kaplan, for example, draws a three-way distinction between the ‘character’ (linguistic meaning), content (roughly, sense) and referent of an indexical expression such as ‘my mother’ (1970). Kaplan’s view may not be inevitable, however. It may be that ‘I’ has an ‘incomplete sense’, in the way that the sense of a functional expression ‘the square root of *x*’ or ‘the author of *x*’ is incomplete. If so, that would explain why the linguistic meaning of ‘I’ cannot be combined with a predicate to yield a complete proposition and there would be no need to postulate a kind of meaning in addition to sense and reference.

Some subscribe to the principle that for one to understand an ‘object-dependent’ proposition - a proposition that such and such an object has such and such a property - one must be able to identify the object in question; one must be able to say who or what it is. This principle seems to be directly contradicted by propositions involving indexicals. I can know that I am hungry or that it is raining now without having any substantive information about who I am or what time it is. Similarly, if I point to something and say ‘I don’t have the slightest idea what that is’ I might well speak the truth! Yet these propositions are object-dependent in that they would not be available if the objects they are about did not exist. Let us mark the fact that speakers can use indexicals with total understanding without having any substantive knowledge of their referents by saying that indexicals are ‘transparent’.

Some indexical sentences have the property of being true in any context of use. For example, one who says (or thinks) ‘I exist’ or ‘I am here now’ or ‘I am saying (or thinking) something now’ must be right. These sentences, then, are somewhat similar to logical truths. But if I say ‘I exist’, then what I say (that HD exists) is not a *necessary* truth, for there are possible circumstances in which I do not exist. So a sentence such as ‘I exist’ is contingent in one respect and yet non-contingent in another. Any theory of indexicality must explain how this is possible.

Perhaps the most puzzling aspect of indexicals is that some of them seem to be ineliminable and their ineliminability seems to have to do with the baffling distinction between subjectivity and objectivity (see Content, indexical). Perry gives the following example (1979). Suppose a meeting is set for Tuesday at noon. All week you believe that the meeting will start at noon on Tuesday. At a certain point, you leave to go to the meeting. Obviously, you must have acquired some new information, a new belief, distinct from the belief that the meeting starts at noon on Tuesday, for you now act in a way you have not acted all week (that is, you leave to go to the meeting). What belief is this? It is the belief that the meeting starts *now*, or, more elaborately, the belief that the meeting starts at noon on Tuesday and it is *now* noon on Tuesday. The use of ‘now’ in describing your new belief seems essential, and not eliminable in favour of non-indexical designations of noon on Tuesday. Moreover, your new ‘indexical belief’ places you within the fabric of time in a way that the beliefs you acquire when you look at your appointment book to determine the time of the meeting do not. An observer ‘outside of time’ could have the belief that the meeting starts at noon on Tuesday, but not the belief that the meeting starts ‘now’. In this way, your indexical belief seems linked to your subjective place in the world. Perry calls such beliefs ‘self-locating’ beliefs. Yet all this is quite puzzling. Is your belief that it is ‘now’ noon on Tuesday something that you can believe only ‘now’? Certainly, you can believe truly that it is now noon on Tuesday only at noon on Tuesday. But can you not at a later time continue to hold precisely the belief you expressed earlier by saying ‘It is now noon on Tuesday’? Yet if you believe at the later time exactly what you believed at the earlier, then why do you not, at the later time, do just what you did at the earlier time - namely, head off to the meeting? Similarly, I believe that I am HD. Can you not believe that too? If you believe just what I do when I believe that I am HD - that is, if you believe exactly the proposition I do, it would appear to follow that our mental states *vis-à-vis* this belief must be identical, then why do you not have any inclination, as I do, to identify yourself as HD? Let us call this ‘the puzzle about

indexical belief'. The puzzle has two parts: first, one argues on the basis of behavioural changes (for example, your getting up to go to the meeting) that an indexical belief (assertion, thought) is not equivalent to any non-indexical counterpart of it. Second, however, similar behavioural changes - or the lack of any - seem to show that such beliefs are 'isolated' - that they cannot be retained over time or shared by different persons: only I can know that I am HD, and that seems counterintuitive. Any adequate theory of indexicality must solve this puzzle about indexical belief.

2 The Kaplan-Perry theory

The most influential theory that has had some success explaining the semantic data about indexicals as outlined above was developed by Kaplan in a series of papers originating in the late 1960s and early 1970s. In the mid-1970s, Perry applied Kaplan's theory to explain the data about the ineliminability of indexicality and the puzzle about indexical belief - data that were first noticed and systematically studied by Castañeda (1967).

According to Kaplan-Perry theory (KPT), each indexical expression has a 'character' or linguistic meaning (Perry's term is 'role') and, given a context, a 'content' in that context. The content in a context of a (declarative) sentence containing an indexical is a 'singular proposition' - a proposition containing the very referent of the indexical as a 'constituent' rather than a sense or concept of that referent. So the content of my use of 'I am hungry now' at a time t is a proposition to the effect that HD is hungry at time t - with HD right there 'in' the proposition, in at least a metaphorical sense. Kaplan takes the character of an indexical to be a function (in the mathematical sense) that takes context into content. Contexts are treated as n -tuples $c = \langle a, t, p, w, \dots \rangle$ with coordinates for the 'agent of the context' (to handle 'I'), the 'time of the context' (to handle 'now'), the 'place of the context', the 'world of the context' and so on (for example, there should be a coordinate for the 'day of the context' to handle 'yesterday', 'today' and 'tomorrow'). To deal with pure demonstratives such as 'this', Kaplan adds a coordinate for a 'demonstration'. This is a kind of abstract counterpart of an act of pointing. For example, the sentence 'That is mine now' will be true at a parameter $\langle a, t, p, w, d, \dots \rangle$ if the demonstrated object d belongs to agent a at t in place p and world w .

Kaplan distinguishes between 'contexts' and 'circumstances'. Formally, the latter are pairs $\langle t, w \rangle$ with t a time and w a possible world. Strictly, expressions are evaluated relative to parameters of the form $\langle \text{context}, \text{circumstance} \rangle$. So the semantic rules are double-indexed. In addition, Kaplan claims that indexicals are 'directly referential'. By this he means that an indexical refers without the aid of a Fregean sense. In this respect indexicals are like proper names. But proper names are, in Kaplan's view, unlike indexicals in that the latter possess characters whereas the former do not (see [Proper names §§2-3](#)).

By appealing to the content/character and context/circumstance distinctions, KPT can readily explain the fact that indexicals are context-dependent, referential and univocal. Moreover, the character/content distinction helps to explain the transparency of indexicals. One need only grasp the character of an indexical to use it with understanding. On occasion, its content may be hidden, as in the case of amnesiacs, who, despite their affliction, can certainly use the first person pronoun perfectly competently. KPT can also explain the special status of sentences such as 'I exist'. The character of 'I exist' always produces a true content in any context. But that content may turn out to be false in further circumstances. For example, if you now write 'I exist' in your diary, the content you express is true now. But, inevitably, that content will be false years hence in circumstances in which you no longer exist and someone then reads these words in your diary. According to KPT, 'I exist' is a logical truth, although its necessitation, 'Necessarily, I exist', and its temporal necessitation, 'It is always the case that I exist', are not. KPT thus involves a novel conception of logical truth. For it is usually assumed that the class of logical truths contains 'Necessarily, P ' and 'It is always the case that P ' if it contains P . Finally, KPT solves the problem of indexical belief by appealing to the notion of character. KPT holds that belief is linked to action via character, not content. If you and I have beliefs with the same content but different character we may well behave differently. But if we have beliefs with the same character and different content we may well behave in the same way. When each of us thinks 'I am about to be attacked by a bear' (same character, different content) we both run for our lives, but when I think 'I am about to be attacked by a bear' and you think 'You are about to be attacked by a bear' (same content, different character) I run for my life and you do not. Similarly, you can retain the belief that the meeting starts now but you can express it at a later time only by using words having a character distinct from that of 'The meeting starts now', and it is the difference in character that accounts for the fact that, at the later

time, you do not do what you did at the earlier time. Likewise, KPT appeals to character to account for the difference between the thought that I am about to be attacked by a bear and the thought that HD is about to be attacked by a bear. The former has a certain character which the latter does not, and that makes all the difference. Similarly, consider the difference between the thought that I am HD and that HD is HD. The former is informative whereas the latter is not. Again, the difference can be traced to the fact that 'I' has a certain character which 'HD' does not.

KPT needs only the character/content and context/circumstance distinctions to account for the semantic data. But it goes further in claiming that indexicals are 'directly referential' and that indexicals give rise to 'singular propositions'. Proponents of KPT argue in several ways for the direct reference thesis. First, Kaplan argues that 'I' does not have the sense of 'the speaker' or any description such as 'the utterer' or 'the thinker'. He would argue, for example, that since 'I am not the speaker' is not contradictory, whereas 'The speaker is not the speaker' is contradictory, 'I' and 'the speaker' are not synonymous. But this argument seems equivocal. Any candidate synonym for 'I' must be context-dependent (and not just parameter-dependent). If we associate a context-dependent rule with 'the speaker' - so that it refers in any circumstance to the 'speaker of the context' - then 'I am not the speaker' is indeed contradictory. It is only by treating 'I' as context-dependent and 'the speaker' as merely parameter-dependent (as it is in one sense) that there is a discrepancy. Second, Kaplan argues in the following way. The descriptive meaning of 'now' is given by the following rule: at *t*, 'now' refers to *t*. Thus, the descriptive meaning of 'now' is 'either inapplicable or irrelevant to determining a referent with respect to a circumstance of evaluation' (1989: 500). Hence, 'now' is directly referential - its descriptive meaning does not serve as the route to its referent in a circumstance. This argument does not work either. One can take the view, as explained above, that the descriptive rule for 'now' is not the one Kaplan assumes. (So the first premise of Kaplan's argument is false.) And if we substitute the double-indexed rule for 'now' it cannot be argued that descriptive meaning (that is, the double-indexed rule) of 'now' is 'irrelevant or inapplicable' to determining its referent in a circumstance.

See also: [Logical and mathematical terms, glossary of](#); [Reference §5](#)

HARRY DEUTSCH

References and further reading

- Braun, D.** (1995) 'What is Character?', *Journal of Philosophical Logic* 24: 227-40. (A critical and relatively nontechnical discussion of some aspects of Kaplan's notion of character.)
- Castañeda, H.-N.** (1967) 'Indicators and Quasi-Indicators', *American Philosophical Quarterly* 4: 85-100. (This is a representative example of several pioneering papers by this author exploring the behaviour of indexicals in propositional attitude constructions.)
- Forbes, G.** (1989) *Languages of Possibility*, Oxford: Oxford University Press. (This work includes a detailed neo-Fregean theory of indexicality. In contrast, the Kaplan-Perry theory discussed above may be described as 'anti-Fregean'.)
- Kamp, H.** (1971) 'Formal Properties of "Now"', *Theoria* 37: 227-73. (This paper develops a double-indexed semantics for certain tense logics with an operator representing 'now'. Kamp was the first to recognize the need for the double-index technique.)
- Kaplan, D.** (1970) 'On the Logic of Demonstratives', *Journal of Philosophical Logic* 8: 81-98. (The pioneering paper in which the key distinctions between character and content and context and circumstance were first introduced.)
- Kaplan, D.** (1989) 'Demonstratives', in *Themes from Kaplan*, ed. J. Almog, J. Perry and H.K. Wettstein, Oxford: Oxford University Press. (This is a classic meditation on the anti-Fregean aspects of the Kaplan-Perry theory.)
- Perry, J.** (1977) 'Frege on Demonstratives', repr. in P. Yourgrau (ed.) *Demonstratives*, Oxford: Oxford University Press, 1990. (This paper attempts to demonstrate the inability of Frege's semantics to handle indexicality and the inadequacy of Frege's own attempts to deal with the phenomenon.)
- Perry, J.** (1979) 'The Problem of the Essential Indexical', *Noûs* 13: 3-21. (By means of several ingenious examples this paper uncovers and attempts to solve the problem of indexical belief.)
- Putnam, H.** (1975) 'The Meaning of "Meaning"', in *Mind, Language and Reality: Philosophical Papers*, vol. 2, Cambridge: Cambridge University Press. (This paper is the main source of the view according to which our terms for natural substances, kinds and phenomena are indexical.)
- Yourgrau, P. (ed.)** (1990) *Demonstratives*, Oxford: Oxford University Press. (This is a collection of essays on

indexicality which emphasize not only semantic issues but also metaphysical and epistemological themes.)

Deontic logic

Deontic logic is the investigation of the logic of normative concepts, especially obligation ('ought', 'should', 'must'), permission ('may') and prohibition ('ought not', 'forbidden'). Deontic logic differs from normative legal theory and ethics in that it does not attempt to determine which principles hold, nor what obligations exist, for any given system. Rather it seeks to develop a formal language that can adequately represent the normative expressions of natural languages, and to regiment such expressions in a logical system.

The theorems of deontic logic specify relationships both among normative concepts (for example, whatever is obligatory is permissible) and between normative and non-normative concepts (for example, whatever is obligatory is possible). Contemporary research beginning with von Wright treats deontic logic as a branch of modal logic, in so far as (as was noted already by medieval logicians) the logical relations between the obligatory, permissible and forbidden to some extent parallel those between the necessary, possible and impossible (concepts treated in 'alethic' modal logic).

1 Standard deontic logic (SDL)

Deontic logic is the logic of normative reasoning. It is common in deontic logic to take one of the three concepts - 'obligatory', 'forbidden', 'permissible' - as primitive and to define the others in terms of the chosen one, the most popular approach defining the permissible and forbidden in terms of the obligatory. In 'standard deontic logic' (SDL) a modal operator O is added to the language of propositional logic so that if p is a sentence of the language, then so is Op , which informally expresses 'It is obligatory that p '. (Alternative approaches treat O as a predicate applied to names of act types, or as an operator on what [Castañeda \(1981\)](#) calls 'practitions'.) The standard definitions of 'permissible' and 'forbidden' are $Pp = \sim O\sim p$ and $Fp = O\sim p$, respectively.

Possible worlds semantics for SDL assumes that relative to any given normative system there is for each possible world w a non-empty set of 'deontic alternatives' for w . This set functions as the normative standard for w , so that Op is said to be true at w (relative to the given normative system) just in case p is true at each of w 's deontic alternatives.

This is put more formally in terms of model structures $\langle W, R \rangle$, where W is a set of possible worlds and R a two-place relation on W . An SDL model is a model structure $\langle W, R \rangle$ plus a function I which assigns truth-values to each $\langle p, w \rangle$, for atomic sentences p and worlds w , and obeys the usual conditions for truth-functional compounds as well as the following condition for O : Op is true at w in $\langle W, R, I \rangle$ just in case p is true in $\langle W, R, I \rangle$ at all worlds w^* such that Rww^* . A sentence p is valid in SDL iff it is true at all worlds of all models.

Just as there are different systems of alethic modal logic (see [Modal logic](#)), different systems of deontic logic are determined semantically by imposing various requirements on R . Consider, for instance, a chief respect in which the parallel between alethic and deontic notions breaks down. 'If it is necessary that p , then p ' is logically valid but, to our daily chagrin, 'If Op , then p ' is not. This difference is reflected in SDL by *not* requiring that R be 'reflexive', that is, not requiring that w be a deontic alternative for itself (whereas an alethic modal system would require reflexivity on its alethic-accessibility relation). For SDL we require only that R is 'serial'; that the set of deontic alternatives for each world is non-empty. It is this requirement that validates $Op \supset \sim O\sim p$. SDL is characterized axiomatically by tautologies T , OT , $Op \supset \sim O\sim p$ and $O(p \supset q) \supset (Op \supset Oq)$. Rules of proof are *modus ponens* and substitution. SDL can be extended by imposing additional constraints on R , such as 'reflexivity once removed' (which validates $O(Op \supset p)$) and 'transitivity' ($Op \supset OOp$), among many others.

A number of so-called 'paradoxes' in SDL motivated the development of alternative systems. The most important inadequacies of SDL were revealed by thought experiments such as [Chisholm's](#) 'contrary to duty imperative' example ([1974](#)). Normative systems often specify what should happen in case certain requirements are violated, for example, that one should apologize or pay reparations or not cry (while mopping up). A set of simple rules of this type may say, for instance, that you ought to tell neither Ann nor Bert a certain secret; and yet were you to go ahead and tell either Ann or Bert then you should tell the other as well. This set obviously is consistent. How can it be represented in SDL? If a abbreviates 'You tell Ann', and b , 'You tell Bert', $O\sim a$ and $O\sim b$ together plausibly represent the primary requirements not to tell either one the secret. What about the conditionals which tell you

what to do if you violate one of these rules? It is here that SDL gets into trouble.

A ‘wide scope’ representation $O(a \supset b)$ of the conditional ‘You ought to tell Bert if you tell Ann’ is a poor choice since it follows in SDL from $O\sim a$ (whereas intuitively the rules are logically independent). Moreover, there is no way to detach expression of the obligation to tell Bert, after you tell Ann: $O(a \supset b)$ and a together do not entail Ob in SDL. (The conditional $a \supset b$ may be true in all deontic alternatives and a true in the actual world, even though b is not true in all the alternatives.) A ‘narrow scope’ representation $a \supset Ob$ of the rule is no better since it follows from $\sim a$ (and yet the rule itself is logically independent of a). Moreover, while this option permits detachment of Ob given a by means of *modus ponens*, Ob is not consistent in SDL with $O\sim b$ (because of the validity of $Oq \supset \sim O\sim q$). So detachment generates an inconsistency that is counterintuitive, since the original rules are logically consistent with the fact that a . This may suggest rewriting the original unconditional obligations as $\sim b \supset O\sim a$ and $\sim a \supset O\sim b$. No inconsistent detachment is possible now, but the suggestion is not viable because the revised rule set would be wholly complied with if you told both Ann and Bert, whereas the rules obviously are violated in some sense if both a and b are true. Theorists seeking formal solutions to these problems proposed two types of response, corresponding broadly to refinements on the wide- and narrow-scope proposals.

2 Two alternatives to SDL

Von Wright proposed a new type of dyadic (two-place) deontic operator to represent expressions of conditional obligation. $O(q/p)$ is read informally as ‘It ought to be that q given p ’. Danielsson, Hansson, Lewis (1973), van Fraassen, Åqvist and others developed semantics in terms of rankings $\leq$ of worlds according to their relative acceptability (similar to the way in which orderings of worlds are used to interpret subjunctive conditionals). Let $\leq$ be a dyadic relation, on a set of worlds, which is transitive and strongly connected (that is, for any two worlds u and v in the set, either $u \leq v$ or $v \leq u$). $O(q/p)$ is true relative to $\leq$ iff some $p \ \& \ q$ -world is ranked more highly in $\leq$ than any $p \ \& \ \sim q$ -world. The unconditional Oq is defined as $O(q/T)$, for tautology T , so that Oq holds just in case q is true in all most acceptable worlds. The rules in the Ann/Bert example may be represented consistently as $\{O\sim a, O\sim b, O(a/b), O(b/a)\}$ and, significantly, it would be consistent to add further rules of the form $O(\sim b/a \ \& \ c)$ and $O(b/a \ \& \ c \ \& \ d)$ and so on. No member of the set is entailed by any other member, and all are independent of factual circumstances. So this language is well suited for representing the content of normative systems with rules and principles that have exceptions or are ‘defeasible’ like $O\sim b$ in our example (there are circumstances, such as when a is true, in which $O\sim b$ should not be used to guide action). Since typically the rules of a system or practice are used to justify actions or states of affairs, a shortcoming of this approach taken alone is that there is no means even for expressing the unconditional ‘all-out’ obligations that may hold in non-ideal circumstances. For instance, in circumstances in which the non-ideal a is true obviously one ought (in some sense) to tell Bert because $O(b/a)$, but this obligation is not entailed by $O(b/a)$ and a . All most acceptable a -worlds may be b -worlds, and a may be true in the actual world, but nothing follows from this because the actual world may not be a most acceptable a -world.

A second approach, developed by Montague, Chellas (1984), Thomason (1981), van Eck (1982), Feldman (1986) and others, ties the semantics of O to specific factual circumstances by interpreting O relative to both worlds and times. One way to do this is to index O to a time t , so that ‘ $O_t q$ ’ is read ‘It ought at time t to be that q ’ and is interpreted relative to what is ‘historically necessary’ (or ‘settled’) at t in w (see Tense and temporal logic). The set of deontic alternatives at t in w is the set of most acceptable worlds (according to a given normative system) of those that are compatible with what is settled at t in w ; and $O_t q$ is true at w just in case q is true at all the deontic alternatives at t in w . So ‘all-out’ obligations at t are relative to what is settled at t . Let the alethic ‘ $N_t q$ ’ say ‘ q is settled at t ’ and be true in w just in case q is true at all worlds historically possible at t in w . At times t when neither a nor b is settled at w (as in our example), $O_t \sim a$ and $O_t \sim b$ are true. But if $N_{t^*} a$, for a later time t^* , then $O_{t^*} b$.

What about the ‘rules’ of the system? There is no straightforward way to formalize them on this approach, though it can be done indirectly by quantifying over time: for example, $(t)(\sim N_t a \supset O_t \sim b)$ replaces $O\sim b$, and $(t)(N_t a \supset O_t b)$ replaces $O(b/a)$. This approach works for the example and permits detachment by *modus ponens* of appropriate all-out obligation statements. The paradoxes of material implication reappear, however, so a first difficulty is distinguishing ‘genuine’ from spurious rules. More importantly, the formal representation of ‘You should not tell Bert’ as a conditional is awkward compared with the English.

3 Combining the two approaches

In any case, the elegant treatment of rules in the first approach is perfectly compatible with the second's temporal treatment of all-out obligation, and combining the two approaches also solves the main problem associated with the first. Given a ranking of worlds, all-out obligations at t are determined by restricting the ranking to worlds that are possible at t . If S expresses all and only what is 'settled' at t in w , then $O_t q$ is true at w iff $O(q/S)$ is true according to $\leq$. This means $O_t q$ is true at w just in case q is true in all most acceptable worlds, according to $\leq$, of those that are still accessible at t in w . (An agent-relative O_t can be formulated in terms of the best worlds within an agent's control.) To detach $O_t q$ from $O(q/p)$ when p does not express all that is settled at t , it is useful to express the idea that there is no r such that $N_t r$ and $\sim O(q/p \& r)$. Let $R_t(q, p)$ say as much. This enables us to reason with $O(a/b)$ even if other rules like $\sim O(a/b \& c)$ also hold. The schema $O(q/p) \& N_t p \& R_t(q, p) \supset O_t q$ can be validated in the combined semantics and a family of normative concepts defined precisely in its terms.

There is an 'all-out' (conclusive) reason for q at t (according to a certain normative system) iff $O_t q$ holds (for that normative system), while q is *prima facie* 'obligatory' at t iff there is some p such that $O(q/p)$ and $N_t p$. $O(q/p)$ is 'undercut' at t iff there is some r such that $N_t r$ and $\sim O(q/p \& r)$; and $O(q/p)$ is 'overridden' at t iff there is some r such that $N_t r$ and $O(\sim q/p \& r)$. $O(q/p)$ is 'defeated' at t iff it is either undercut or overridden. A *prima facie* reason for q based on $O(q/p)$ is defeated if $O(q/p)$ is defeated. There is a *prima facie* conflict at t iff there are p, q, r and s such that $N_t(q \& s)$, $O(p/q)$, $O(r/s)$ and $N_t \sim(p \equiv r)$.

The combined system does not solve all of the problems that SDL encounters. In SDL and in the combined approach, Op entails Oq if p entails q . So, for instance, since h : 'You help the man who was robbed' entails r : 'The man was robbed', $O_t h$ entails $O_t r$, a counterintuitive result (hence the 'Good Samaritan paradox').

Also in the combined system, as in SDL, no all-out conflict (for example, of the form $O_t p \& O_t q$, where p and q are contraries) is consistent, but arguably there are coherent normative systems in which genuine conflicts are a possibility, and if so it is not the business of deontic logic to legislate otherwise.

So the combined system, like SDL, is arguably too 'strong' in the sense that it validates too much. It can be weakened semantically by interpreting $O(q/p)$ not in terms of $\leq$, but rather in terms of a 'neighbourhood' selection function f that assigns to each proposition (set of worlds) q a set of subsets of W (the set of propositions that are obligatory given q). $O(q/p)$ is true iff $q \in f(p)$. Let $R_t(q, p)$ be true iff for all r , if $N_t r$, then $q \in f(p \& r)$; and $O_t q$ is true iff there is some p such that $O(q/p)$, $N_t p$ and $R_t(q, p)$. $O_t p$ does not entail $O_t q$ even if p entails q , and $O_t p \& O_t q$ is consistent even for contraries p and q . (In general, there is an 'all-out conflict' at t iff there is a *prima facie* conflict at t , as described above, and neither $O(p/q)$ nor $O(r/s)$ is defeated at t .) There is so little logical structure here (the only rule of proof is substitution) that it is difficult to envisage complaints grounded in its excessive strength. None the less, it can be used, exactly as above, to define all of the concepts defined earlier. While the weakness of this very minimal system probably is excessive, at least for applications to many normative systems, it is an appropriate starting point for developing stronger systems, and for structuring discussion of the appropriate strength of deontic logics for particular applications.

See also: [Logical and mathematical terms, glossary of](#)

MARVIN BELZER

References and further reading

- Castañeda, H.-N.** (1981) 'The Paradoxes of Deontic Logic: The Simplest Solution to All of Them in one Fell Swoop', in R. Hilpinen (ed.) *New Studies in Deontic Logic*, Dordrecht: Reidel, 37-85. (Develops a distinction between propositions and partitions, and uses this distinction to address the problems of SDL in a unique way.)
- Chellas, B.F.** (1980) *Modal Logic*, Cambridge: Cambridge University Press, esp. 190-203, 272-7. (Discusses SDL and its extensions, the dependence of obligation on time, and conditional obligation.)
- Chisholm, R.** (1974) 'Practical Reason and the Logic of Requirement', in S. Korner (ed.) *Practical Reason*, New Haven, CT: Yale University Press, 1-16. (Presents an informal account of many of the definitions in §3.)
- Eck, J. van** (1982) 'A System of Temporally Relative Modal and Deontic Predicate Logic and its Philosophical Applications', *Logique et Analyse* 25: 249-90, 339-81. (Develops a temporally relative logic as in the 'second

approach' with quantifiers and a subjunctive conditional, and offers alternative definitions of some of the terms defined in §3.)

- Feldman, F.** (1986) *Doing the Best We Can*, Dordrecht: Reidel. (Application to an act-utilitarian ethics of a temporally relative deontic logic that uses world rankings and defines an agent-relative conception of 'settled'.)
- Føllesdal, D. and Hilpinen, R.** (1971) 'Deontic Logic: An Introduction', in R. Hilpinen (ed.) *Deontic Logic: Introductory and Systematic Readings*, Dordrecht: Reidel, 1-35. (Good introduction to SDL and its history.)
- Humberstone, I.** (1983) 'The Background of Circumstances', *Pacific Philosophical Quarterly* 64: 19-34. (Develops a time- and agent-relative all-out ought.)
- Lewis, D.K.** (1973) *Counterfactuals*, Oxford: Blackwell. (Discusses the defeasibility of expressions of conditional obligation and presents logics for the dyadic operator discussed in §2.)
- Lewis, D.K.** (1974) 'Semantic Analyses for Dyadic Deontic Logic', in S. Stenlund (ed.) *Logical Theory and Semantic Analysis*, Dordrecht: Reidel, 1-14. (Compares different semantics for the dyadic operator discussed in §2.)
- Loewer, B. and Belzer, M.** (1983) 'Dyadic Deontic Detachment', *Synthese* 54: 295-319. (Expands on the 'combined approach' discussed in §3.)
- Thomason, R.** (1981) 'Deontic Logic as Founded on Tense Logic', in R. Hilpinen (ed.) *New Studies in Deontic Logic*, Dordrecht: Reidel, 165-76. (Discusses the dependence of obligation on time as in the 'second approach'.)

Descriptions

'Definite descriptions' are noun phrases of the form 'the' + noun complex (for example, 'the finest Greek poet', 'the cube of five') or of the form possessive + noun complex (for example, 'Sparta's defeat of Athens'). As Russell realized, it is important to philosophy to be clear about the semantics of such expressions. In the sentence 'Aeschylus fought at Marathon', the function of the subject, 'Aeschylus', is to refer to something; it is a referential noun phrase (or 'singular term'). By contrast, in the sentence 'Every Athenian remembers Marathon', the subject noun phrase, 'every Athenian', is not referential but quantificational. Definite descriptions appear at first sight to be referential. Frege treated them referentially, but Russell held that they should be treated quantificationally in accordance with his theory of descriptions, and argued that certain philosophical puzzles were thereby solved.

1 Frege

Gottlob Frege provided the first systematic account of quantification in natural language and the first systematic theory of reference (1892). The class of 'singular terms' (referential noun phrases), for Frege, was delimited by a set of logical tests (for example, the licensing of existential generalization) and was recursive. It included ordinary proper names and definite descriptions. Thus '5', 'the cube of 5', 'Aeschylus' and so on were all singular terms.

If a description 'the F ' is referential, then it is natural to take its reference to be the unique entity satisfying ' F '; a sentence 'The F is G ' is true if and only if that entity is G . But what if no entity, or more than one entity, satisfies the descriptive condition, as in (1) or (2)?

- (1) The largest prime number lies between 10^{23} and 10^{27} .
- (2) The man who landed on the moon was American.

Such descriptions are said to be 'improper'. Frege considered it a defect of natural language that it permits the possibility of improper terms. As far as his own logical system was concerned, he thought it essential that every formula have a truth-value, and so he insisted that every singular term have a reference (or meaning): he stipulated that a specified object in the range of the quantifier(s) serve as the referent of every improper description. While this stipulation proved useful for his formal system, Frege recognized that some other account of improper descriptions in ordinary language was needed. Once he had made his distinction between the 'sense' and 'reference' of an expression, he suggested that an improper term has a sense but no reference (see Frege, G. §3; *Sense and reference* §1). The main problem with this proposal is that it predicts, rather counterintuitively, that any sentence containing an improper term (in a transparent context) lacks a truth-value. For example, it predicts that (1) and (2) above lack truth-values.

2 Russell's theory of descriptions: informal characterization

Like Frege, Bertrand Russell thought it important to explain how a sentence such as (1) or (2) could be meaningful. At one time he entertained the idea of a realm of non-existent entities to serve as the referents of descriptions such as 'the largest prime number' and 'the round square'; but by 1905 he thought this idea conflicted with a 'robust sense of reality' and his theory of descriptions came about, in part as an attempt to purify his ontology (see *Existence* §2).

On Russell's account, descriptions are not singular terms at all but phrases that logical analysis reveals to be quantificational: if 'the F ' is a definite description and ' $\dots$ is G ' is a predicate phrase, then the proposition expressed by an utterance of 'The F is G ' is equivalent, says Russell, to the proposition expressed by an utterance of 'There is exactly one F , and everything that is F is G '. That is, 'The F is G ' is analysed as

- (3) $(\exists x)((\forall y)(Fy \leftrightarrow y = x) \ \& \ Gx)$.

The proposition expressed by 'The F is G ' is 'general' ('object-independent') rather than 'singular' ('object-dependent') in the sense that there is no object for which its grammatical subject stands; upon whose existence that of the proposition expressed depends. Unlike a singular term, a definite description, even if it is in fact satisfied by a unique object, does not actually *refer* to that object. It is as wrong, on Russell's account, to inquire into the referent of 'the F ' as it is to inquire into the referent of 'every F ' or 'no F '.

On Russell's account, sentences containing improper descriptions have truth-values. For example, (1) above is false as it is not the case that there exists a largest prime number. Similarly, sentence (2) is false as it is not the case that there exists exactly one man who landed on the moon.

Russell's theory opens up the possibility of accounting for certain *de dicto/de re* ambiguities in terms of scope permutations (see *De re/de dicto*). For example, (4) below may be analysed as either (5) or (6), according to whether the description 'the largest prime number' is given large or small scope with respect to 'John thinks that':

(4) John thinks that the largest prime number lies between 10^{23} and 10^{27} .

(5) $(\exists x)((\forall y)(\text{largest-prime } y \leftrightarrow y = x) \ \& \ \text{John thinks that: } x \text{ lies between } 10^{23} \text{ and } 10^{27})$.

(6) John thinks that: $(\exists x)((\forall y)(\text{largest-prime } y \leftrightarrow y = x) \ \& \ x \text{ lies between } 10^{23} \text{ and } 10^{27})$.

(5) is false; but (6) may be true. Thus Russell is able to explain the intuitive ambiguity of (4) and avoid positing an ontology that includes such things as a largest prime number. Smullyan (1948) points out that Russell's theory similarly explains *de dicto/de re* ambiguities in modal contexts, for example, in 'The number of planets is necessarily odd' (see *De re/de dicto* §2; *Modal logic*).

Russell came to treat ordinary proper names as 'disguised' or 'truncated' descriptions. For example, the name 'Cicero' might be analysed as the description 'the greatest Roman orator' while the coreferential name 'Tully' might be analysed as the description 'the author of *De Fato*'. On the face of it, this provided Russell with accounts (not dissimilar to Frege's) of why 'Cicero was bald' and 'Tully was bald' differ in informativeness, and of why (7) and (8) need not agree in truth-value:

(7) John believes Cicero was bald.

(8) John believes Tully was bald.

In the light of Kripke's work on names and necessity (1980), it is widely held that descriptive analyses of proper names cannot succeed (see *Proper names*). There is good reason, however, to think that at least some pronouns anaphorically linked to (referring back to), but not bound by, quantified noun phrases are understood in terms of definite descriptions (see Neale 1990).

3 Russell's theory of descriptions: formal characterization

On Russell's account, descriptions are 'incomplete' symbols; they have 'no meaning in isolation', that is, they do not stand for things. In *Principia Mathematica* (1910-13), descriptions are represented by quasi-singular terms of the form ' $(\iota x)(Fx)$ ', which can be read as 'the unique x which is F '. Superficially, the iota-operator is a variable-binding device for forming a term from a formula. A predicate symbol ' G ' may be prefixed to a description ' $(\iota x)(Fx)$ ' to form a formula ' $G(\iota x)(Fx)$ ', which can be expanded in accordance with a suitable 'contextual definition'. (To define an expression ζ contextually is to provide a procedure for converting any sentence containing occurrences of ζ into an equivalent sentence that is ζ -free.)

The analysis in (3) above does not constitute a final contextual definition of ' $G(\iota x)(Fx)$ ' because of the possibility of scope ambiguity where a formula containing a description is itself a constituent of a larger formula (see *Scope*). Scope ambiguity is conveniently illustrated with descriptions in the context of negation. For a genuine singular term α , there is no difference between wide and narrow scope negation: α is not- F just in case it is not the case that α is F . For a description, however, there is a formal ambiguity. Let ' Kx ' represent ' x is a king of France' and ' Wx ' represent ' x is wise'. Then the formula ' $\neg W(\iota x)(Kx)$ ' ('The king of France is not wise') is ambiguous between (9) and (10):

(9) $(\exists x)((\forall y)(Ky \leftrightarrow y = x) \ \& \ \neg Wx)$

(10) $\neg(\exists x)((\forall y)(Ky \leftrightarrow y = x) \ \& \ Wx)$.

These are not equivalent: only (10) can be true when there is no king of France. In *Principia Mathematica*, the scope of a description is specified by appending a copy of it within square brackets to the front of the formula that constitutes its scope. Thus (9) and (10) are represented as (11) and (12) respectively:

(11) $[(\iota x)(Kx)]\neg\{W(\iota x)(Kx)\}$

(12) $\neg\{[(\iota x)(Kx)]W(\iota x)(Kx)\}$.

In (11) the description has what Russell calls a ‘primary occurrence’ by virtue of having scope over the negation; in (12) the description has a ‘secondary occurrence’ by virtue of lying within the scope of the negation. Where a description has smallest possible scope, it is conventional to omit the scope marker; thus (12) can be reduced to ‘ $\neg W(\iota x)(Kx)$ ’.

With the matter of scope behind us, the theory of descriptions can be stated exactly:

$$[(\iota x)(\phi x)]G(\iota x)(\phi x) =_{\text{df}} (\exists x)((\forall y)(\phi y \leftrightarrow y = x) \ \& \ Gx),$$

where ϕ is a formula. On Russell’s account, there is no possibility of a genuine referring expression failing to refer, so no predicate letter in the language of *Principia Mathematica* stands for ‘exists’. Russell introduces a symbol ‘E!’ (‘E shriek’) that may be combined with a description to create a well-formed formula. Thus

$$E!(\iota x)(\phi x) =_{\text{df}} (\exists x)(\forall y)(\phi y \leftrightarrow y = x).$$

‘E!’ allows a treatment of negative existentials. (According to Russell, an utterance of ‘The king of France does not exist’ made today would be true precisely because there is no king of France.) Successive applications will allow any well-formed formula containing a definite description to be replaced by an equivalent formula that is description-free.

It is often objected that Russell’s theory, which substitutes complex quantificational structure for ‘the’, is unfaithful to surface syntax. The objection is engendered by an insufficiently keen appreciation of the distinction between a theory and its formal implementation. The extent of the mismatch between ‘The king is wise’ and its analysis

$$(13) (\exists x)((\forall y)(\text{king } y \leftrightarrow y = x) \ \& \ \text{wise } x)$$

has nothing to do with descriptions *per se*. In order to characterize the logical forms of even ‘some philosophers are wise’ and ‘every philosopher is wise’ in the predicate calculus we have to use formulas containing sentence connectives, no counterparts of which occur in the surface forms of the sentences:

$$(14) (\exists x)(\text{philosopher } x \ \& \ \text{wise } x)$$

$$(15) (\forall x)(\text{philosopher } x \rightarrow \text{wise } x).$$

And when we formalize sentences such as ‘Just two philosophers are wise’, we find much more complexity than there is in surface syntax:

$$(16) (\exists x)(\exists y)[\text{philosopher } x \ \& \ \text{philosopher } y \ \& \ \text{wise } x \ \& \ \text{wise } y \ \& \ (\forall z)((\text{philosopher } z \ \& \ \text{wise } z) \rightarrow z = x \vee z = y)].$$

The supposed problem about descriptions, then, is in fact a symptom of a larger problem involving the application of first-order logic to sentences of ordinary language.

Work on ‘generalized’ quantification provides a solution to the larger problem (as well as treatments of quantifiers such as ‘most’ that cannot be handled within first-order logic; see [Quantifiers, generalized](#)). Natural language quantification is normally restricted: we talk about all philosophers or most poets, not about all or most entities. A simple modification of the predicate calculus yields a language - call it ‘RQ’ - that captures this fact while retaining the precision of regular first-order logic. In RQ, a determiner such as ‘some’, ‘every’ or ‘no’ combines with a formula to create a restricted quantifier such as ‘[every x : philosopher x]’. And such a quantifier may combine with a formula to form a formula:

$$(17) [\text{every } x: \text{philosopher } x] (\text{wise } x).$$

The viability of such a language shows that the language of *Principia Mathematica* is not essential to the theory of descriptions. Since the word ‘the’ is a one-place quantificational determiner (as are ‘some’, ‘every’, ‘no’ and so on), RQ can treat ‘the’ as combining with a formula ‘king x ’ to form a restricted quantifier ‘[the x : king x]’. The sentence ‘The king is wise’ will then be represented as

$$(18) [\text{the } x: \text{king } x] (\text{wise } x).$$

Different scope possibilities are easily captured. For instance, ‘The king is not wise’ is ambiguous between (19) and (20):

(19) [the x : king x] $\neg$ (wise x)

(20) $\neg$ [the x : king x](wise x).

Using a formal language in which descriptions are treated as restricted quantifiers does not mean abandoning Russell’s view that descriptions are ‘incomplete symbols’ that ‘disappear on analysis’. Rather, treating descriptions as restricted quantifiers results in an explanation of where his theory of descriptions fits into a systematic account of natural language quantification, a theory in which ‘every’, ‘some’, ‘most’, ‘a’, ‘the’ and so on are members of a unified syntactic and semantic category.

4 Strawson’s theory and criticisms of Russell

As part of a broad critique of the idea that the semantics of formal languages can be used to analyse the meanings of statements of natural language, **P.F. Strawson** argued against Russell’s theory of descriptions on the grounds that (1) it fails to recognize that referring is something done by speakers and not expressions, (2) it fails to do justice to the way speakers ordinarily use sentences containing descriptions to make statements (speakers use descriptions to refer, not to quantify) and (3) it rides roughshod over important distinctions, such as the distinction between the meaning of a sentence σ and the statement made by a particular use of σ (see **Propositions, sentences and statements**).

Using as an example ‘The present king of France is wise’, Strawson argues that Russell’s theory is thwarted because the same sentence can be used to say something true on one occasion and something false on another. It is certainly true that Russell paid little attention to the distinction between the linguistic meaning of a sentence type and the proposition expressed by a particular dated utterance of that sentence type; but it was the latter that actually concerned him, and Strawson could get no mileage out of Russell’s inattention to the distinction. The fact that a description (or any other quantified noun phrase) may contain an indexical component (‘the *present* king of France’, ‘every man *here*’, and so on) illustrates that some descriptions are subject to both the theory of descriptions and a theory of indexicality (see **Demonstratives and indexicals**). Thus contextual features play a role in fixing the proposition expressed. And this can be true also if the overt indexical element is missing, as in ‘The king of France is wise’.

This appreciation of contextual factors forms the basis of the Russellian response to a second Strawsonian objection. According to this, someone who uses a description ‘the F ’ typically intends to refer to some object or other and say something about it; there is no question of claiming that some object uniquely satisfies F . Someone who says ‘The table is covered with books’, for instance, does not express a proposition that entails the existence of exactly one table. But, Strawson claimed, it is a part of the meaning of ‘the F ’ that such an expression is used correctly only if there is an F . If this condition is not satisfied - if the ‘presupposition’ that there is an F is false - a use of ‘The F is G ’ cannot be considered to express a proposition that is either true or false. The Russellian response to Strawson is that descriptions such as ‘the table’ are often understood as elliptical uses of fuller descriptions such as ‘the table over there’, ‘the table in front of me’ and so on; or else they are evaluated with respect to a restricted domain of discourse. Again the phenomenon is not confined to descriptions, but is found with quantified noun phrases more generally.

Strawson’s original statement (1950) of his own theory contains an interesting ambiguity. He can be understood as claiming either that no proposition is expressed, or that a proposition which is neither true nor false is expressed, when someone uses a sentence containing an empty description. A second ambiguity comes with the notion of ‘presupposition’. This can be viewed as an epistemological or pragmatic relation between a person and a statement, or as a logical relation between two statements (see **Presupposition**). An epistemological or pragmatic notion of presupposition appears to have no bearing on the semantic issues Strawson wanted to address when he challenged Russell.

The Strawsonian position faces some serious obstacles. If someone were to utter (21) right now, they would unquestionably say something false.

(21) The king of France shot my cat last night.

But on Strawson's account, the speaker will have expressed no proposition because the presupposition that there is a unique king of France is false. Descriptions occurring in the context of attitude verbs create a similar problem. For example, someone might utter a true statement using (22):

(22) Ponce de León thought the fountain of youth was in Florida,

so the presence of an empty description does not always result in a failed speech act. This is something Strawson (1964) came to concede. In order to reduce the number of incorrect predictions made by his earlier theory, he suggested that the presence of an empty description sometimes renders the proposition expressed false and at other times prevents a proposition from being expressed at all (sometimes 'The *F* is *G*' entails the existence of a unique *F*, and at other times it (only) presupposes it).

5 Ambiguity theories

Consideration of the behaviour of descriptions in non-extensional contexts and the possibility of misdescribing an individual as the *F*, but successfully communicating something about that individual, has led some philosophers (for example, Donnellan 1966) to suggest that descriptions are sometimes quantificational, at other times referential. When 'the *F*' is used in the Russellian way, the proposition expressed is general (object-independent); when it is used referentially the proposition expressed is singular (object-dependent). Donnellan considers examples such as the following: (1) A detective discovers Smith's mutilated body but has no idea who killed him. Looking at the body, he exclaims, 'The murderer is insane'. (2) Jones is on trial for Smith's murder; I am convinced of his guilt; hearing Jones ranting in court, I say, 'The murderer is insane'. On Donnellan's account, in case (1) the description is being used attributively; in case (2) it is being used referentially. Cases such as (2), it is argued, cannot be treated in accordance with Russell's theory. Following Grice (1969), however, many have argued (1) that so-called referential uses of descriptions can usually be accommodated within Russell's theory by invoking a distinction between the proposition expressed by (an utterance of) a sentence on a given occasion and the proposition the speaker primarily intends to communicate on that occasion; (2) that the phenomenon of referential usage is not specific to definite descriptions, but arises with quantified noun phrases quite generally; (3) that the referential/attributive distinction is neither exclusive nor exhaustive; and (4) that no such distinction can do the work of Russell's notion of the scope of a description. It would seem, then, that something very close to Russell's theory will probably form a component of any finally acceptable theory.

See also: [Free logics, philosophical issues in §3](#); [Logical and mathematical terms, glossary of](#); [Reference §6](#)

STEPHEN NEALE

References and further reading

- Carnap, R.** (1947) *Meaning and Necessity*, Chicago, IL: University of Chicago Press. (Discussion of views presented in §§1-2.)
- Davies, M.** (1981) *Meaning, Quantification and Necessity*, London: Routledge. (Discussion of views presented in §5.)
- Donnellan, K.** (1966) 'Reference and Definite Descriptions', *Philosophical Review* 77: 281-304. (Discussion of views presented in §5.)
- Frege, G.** (1892) 'Über Sinn und Bedeutung', *Zeitschrift für Philosophie und philosophische Kritik* 100: 25-50; trans. 'On Sense and Reference', in *Translations from the Philosophical Writings of Gottlob Frege*, trans. and ed. P.T. Geach and M. Black, Oxford: Blackwell, 3rd edn, 1980. (Source of views presented in §1. The page numbers of the original publications appear in the text.)
- Grice, H.P.** (1969) 'Vacuous Names', in D. Davidson and J. Hintikka (eds) *Words and Objections: Essays on the Work of W.V. Quine*, Dordrecht: Reidel, 118-45. (Discussion of views presented in §§2, 5.)
- Kaplan, D.** (1978) 'Dthat', in P. Cole (ed.) *Syntax and Semantics*, vol. 9, *Pragmatics*, New York: Academic Press, 221-43. (Discussion of views presented in §5.)
- Kripke, S.A.** (1979) 'Speaker Reference and Semantic Reference', in P.A. French, T.E. Uehling and H.K. Wettstein (eds) *Contemporary Perspectives in the Philosophy of Language*, Minneapolis, MN: University of Minnesota Press, 6-27. (Discussion of views presented in §§2, 5.)
- Kripke, S.A.** (1980) *Naming and Necessity*, Cambridge, MA: Harvard University Press. (See §2.)

- Neale, S.** (1990) *Descriptions*, Cambridge, MA: MIT Press. (Discussion of views presented in §§2-5. Extensive bibliography.)
- Neale, S.** (1994) 'Grammatical Form, Logical Form, and Incomplete Symbols', in A.D. Irvine and G.A. Wedeking (eds) *Russell and Analytic Philosophy*, Toronto, Ont.: University of Toronto Press, 97-139. (Discussion of views presented in §2.)
- Russell, B.A.W.** (1905) 'On Denoting', *Mind* 14: 479-93; repr. in *Logic and Knowledge: Essays 1901-1950*, ed. R.C. Marsh, London: Routledge, 1992. (Source of views presented in §2.)
- Russell, B.A.W.** (1911) 'Knowledge by Acquaintance and Knowledge by Description', in *Mysticism and Logic*, London: Routledge, 1986. (Source of views presented in §2.)
- Sainsbury, R.M.** (1979) *Russell*, London: Routledge & Kegan Paul. (Discussion of views presented in §§2, 5.)
- Sellars, W.S.** (1954) 'Presupposing', *Philosophical Review* 63: 197-215. (Discussion of views presented in §4.)
- Smullyan, A.F.** (1948) 'Modality and Description', *Journal of Symbolic Logic* 13: 31-7; repr. in L. Linsky (ed.) *Reference and Modality*, Oxford: Oxford University Press, 1971, 35-43. (Discussion of views presented in §2.)
- Strawson, P.F.** (1950) 'On Referring', *Mind* 59: 320-44. (Source of views presented in §4.)
- Strawson, P.F.** (1954) 'Reply to Mr. Sellars', *Philosophical Review* 63: 216-31. (Source of views presented in §4.)
- Strawson, P.F.** (1964) 'Identifying Reference and Truth-Values', *Theoria* 30: 96-118. (Source of views presented in §4.)
- Strawson, P.F.** (1972) *Subject and Predicate in Logic and Grammar*, London: Methuen. (Source of views presented in §4.)
- Whitehead, A.N. and Russell, B.A.W.** (1910-13) *Principia Mathematica*, Cambridge: Cambridge University Press, 3 vols; 2nd edn, 1925-7, esp. vol. 1, * 14. (Source of views presented in §3.)

Dialogical logic

Dialogical logic characterizes logical constants (such as ‘and’, ‘or’, ‘for all’) by their use in a critical dialogue between two parties: a proponent who has asserted a thesis and an opponent who challenges it. For each logical constant, a rule specifies how to challenge a statement that displays the corresponding logical form, and how to respond to such a challenge. These rules are incorporated into systems of regimented dialogue that are games in the game-theoretical sense. Dialogical concepts of logical consequence can then be based upon the concept of a winning strategy in a (formal) dialogue game: B is a logical consequence of A if and only if there is a winning strategy for the proponent of B against any opponent who is willing to concede A. But it should be stressed that there are several plausible (and non-equivalent) ways to draw up the rules.

1 Main characteristics

Logic is to provide us with viable concepts of logical consequence, logical truth, consistency and so on, and to that end to elucidate the meaning of logical constants (see **Logical constants**). Dialogical logic, which is the logical part of dialogue theory, goes about this by assigning to each logical constant a clear ‘meaning-in-use’, where the context of use is one of critical dialogue arising from a conflict of opinions about a statement: the initial thesis. This thesis - the question at issue, one might say - is challenged by the opponent (O) and defended by the proponent (P). The opponent may or may not have granted some statements in advance: the initial concessions or hypotheses of the dialogue. The dialogical approach is to be compared to, and contrasted with, semantic or model-theoretic characterizations of the logical constants (see **Model theory**) and deduction-theoretic characterizations (see **Proof theory**; **Natural deduction**, **tableau** and **sequent systems**).

Each logical constant is characterized by the specific modes of challenge of which the participants may avail themselves to attack or question a statement with that particular logical constant as its principal operator; as well as by the specific responses to these challenges. For instance, a conjunctive statement ‘A and B’ can be challenged in two ways: by questioning A and by questioning B. The response to the first type of challenge consists of A, whereas the response to the second type of challenge consists of B. In the case of a disjunctive statement ‘A or B’, however, the only type of challenge pertains to this statement in its totality, but the responses are the same, with the distinctive difference that in this case the choice between A and B is the respondent’s, rather than the challenger’s. A survey of these so-called logical rules is given in the table.

The logical rules are not sufficient, for the meaning of each logical constant also depends on global characteristics of the dialogues. Structural rules stipulate who may challenge what and how often, whether responses may be repeated, and so on. Finally, rules for winning and losing determine the situations in which a participant may be declared to be the winner (or loser) of the dialogue. A full set of rules of all three types (see the table and lists of rules below) defines a ‘dialogue game’, or ‘dialectic system’.

If a dialogue game makes use of an interpreted nonlogical vocabulary, the game and its dialogues are called material; otherwise they are called formal. In material games, typically, elementary statements can be defended by pointing out their (extra-dialogical) truth. Thus observation and other ways of fact-finding are connected with material dialogues. In formal dialogue games these material modes of defence are not available.

2 An example

The following formal dialogue game yields intuitionistic (constructive, effective) predicate logic (see **Intuitionism**). It is based on a first-order language for predicate logic with an infinity of individual constants $a, b, \dots$. Formulas are denoted as $A, B, \dots$. The result of substituting an individual constant a for the free variable x in the formula A is denoted as $A[a/x]$.

This game is asymmetric in the following sense: at all times P has a statement to defend, whereas O never has a thesis to defend. However, P must have some means to exploit the initial concessions. To that end P is allowed to question the concessions in order to generate more detailed concessions. The corresponding moves of question (by P) and answer (by O) are formally identical to the moves of challenge (properly so called) by O and response (or, defence) by P , but the intuitive motivation is different. Thus the asymmetry of the game remains hidden as far as the logical rules are concerned. It is clearly borne out, however, by the structural rules and by the very different

ways P and O can win a dialogue.

Logical Rules

	Statement by proponent	Challenge/question by opponent	Defence/answer by proponent
Rule $\rightarrow$	$A \rightarrow B$	A	B
Rule $\neg$	$\neg A$	A	$\perp$ (elementary statement of absurdity)
Rule $\vee$	$A \vee B$	?	A or B (proponent chooses)
Rule $\wedge$	$A \wedge B$	A? or B? (opponent chooses)	A B
Rule $\forall$	$\forall x A$	a? (opponent selects a constant)	$A[a/x]$
Rule $\exists$	$\exists x A$	?	$A[a/x]$ (proponent selects a constant)
Rule E_l	elementary statement	?	none

Structural Rules

- 1 There are two participants: the proponent, P , who defends a thesis, and the opponent, O , who may or may not have granted a number of concessions.
- 2 The participants move alternately. O makes the first move: a challenge of the thesis.
- 3 Each move by O is a challenge or an answer according to a logical rule.
- 4 Each move by P is a question or a defence according to a logical rule, or a winning remark, that is, either '*Ipse dixisti!*' ('You said so yourself!') or '*Absurdum dixisti!*' ('You said something absurd!').
- 5 P is not allowed to question O 's elementary statements.
- 6 A winning remark '*Ipse dixisti!*' is allowed only if the most recently challenged statement of P 's can also be found among O 's concessions.
- 7 A winning remark '*Absurdum dixisti!*' is allowed only if $\perp$ can be found among O 's concessions.
- 8 The only statement P is allowed to defend is the one that was most recently attacked by O .
- 9 Except for the initial challenge, each move by O is to consist of a reaction to the immediately preceding move by P (answering P 's question, or challenging a newly introduced statement).
- 10 Before the onset of the dialogue proper, P is to announce a limit on the number of P -moves. P is not allowed any moves beyond that number.

Rules for Winning and Losing

- 1 P wins by making a winning remark.
- 2 O wins whenever it is P 's turn to make a move but no legal move is available.

3 Motivations and applications

The connection between dialogues and logic, which goes back to Aristotle's *Topics*, was reintroduced by Paul Lorenzen in 1958 (see [Lorenzen and Lorenz 1978](#)). Originally, the motivation was to be found in the philosophy of mathematics: the dialogues were to yield a criterion of constructivity and to justify a constructive logic without

recourse to Brouwerian solipsism (see [Intuitionism](#)). But studies by Kuno Lorenz and others were not slow to point out that with a little fiddling of the structural rules the dialogue games could be forced to yield classical logic instead, although constructive logic seemed more ‘natural’ from the dialogical point of view.

In the meantime it had become clear how logical concepts were to be defined dialogically. We have seen above how a definition of logical consequence can be framed in terms of the existence of a winning strategy for the proponent in a formal game. The case of an empty set of concessions gives us a definition of logical truth. Consistency of a set of statements can be defined as the existence of a winning strategy for an opponent who concedes these statements against a proponent who defends $\perp$. This strategy shows how *O* can maintain this position without conceding absurdity. Clearly, different formal dialogue games may give rise to different concepts of logical consequence; to different logics, one might say.

The challenges and responses given in the logical rules of dialogical logic are closely similar and often identical to those given in Hintikka’s rules for game-theoretic semantics (see [Semantics, game-theoretic](#)). But in game-theoretic semantics one tries to exploit the concept of a game to construct a semantic theory (for natural language), whereas dialogical logic is rather to be conceived as providing an alternative to the semantic approach. Even so, the two approaches are closely related.

Later on dialogical logic came to be part of a programme for a normative reconstruction of the language of science, ethics, philosophy and politics that was worked out by Paul Lorenzen and others see [Lorenzen and Schwemmer 1973](#)) in the so-called Erlangen school of German constructivism (see [Lorenzen, P.](#)). The dialogue rules contribute to this programme by giving us reconstructed versions of the logical constants.

More generally, dialogical logic has been influential in the development of argumentation theory, including the theory of argumentative discussions and fallacy theory.

ERIK C.W. KRABBE

References and further reading

- Barth, E.M. and Krabbe, E.C.W.** (1982) *From Axiom to Dialogue: A Philosophical Study of Logics and Argumentation*, Berlin and New York: De Gruyter. (Chapters 3 and 4 provide normative foundations, from the point of view of a theory of argumentative discussions, for a number of dialectic systems. Chapters 5, 7 and 11 discuss the metatheory of dialectic systems and present proofs of equivalence with model-theoretic and deduction-theoretic systems. A standard introductory course in logic is presupposed.)
- Felscher, W.** (1986) ‘Dialogues as a Foundation for Intuitionistic Logic’, in D. Gabbay and F. Guenther (eds) *Handbook of Philosophical Logic*, vol. 3, *Alternatives to Classical Logic*, Dordrecht: Reidel, 341-72. (A useful and concise introduction to the subject, with a useful bibliography. A bit more technical.)
- Haas, G.** (1984) *Konstruktive Einführung in die formale Logik (A Constructive Introduction to Formal Logic)*, Mannheim: Bibliographisches Institut, Wissenschaftsverlag. (This is both an introduction to logic and an introduction to dialogical logic, with many examples, detailed explanations and proofs.)
- Krabbe, E.C.W.** (1985) ‘Formal Systems of Dialogue Rules’, *Synthese* 63 (3): 295-328. (Includes a survey of options in constructing a dialectic system, with a closer study of constructive systems and equivalence proofs. Only the last part is highly technical. Useful bibliography.)
- Lorenzen, P.** (1969) *Normative Logic and Ethics*, Mannheim: Bibliographisches Institut. (An early and brief exposition in English.)
- Lorenzen, P.** (1986) *Lehrbuch der konstruktiven Wissenschaftstheorie (Textbook of Constructive Philosophy of Science)*, Mannheim: Bibliographisches Institut, Wissenschaftsverlag. (Contains the first part of [\(Lorenzen and Schwemmer \(1973\).\)](#)
- Lorenzen, P. and Lorenz, K.** (1978) *Dialogische Logik*, Darmstadt: Wissenschaftliche Buchgesellschaft. (The essential writings of Lorenzen and Lorenz from the genesis of contemporary dialogical logic in 1958 to 1973.)
- Lorenzen, P. and Schwemmer, O.** (1973) *Konstruktive Logik, Ethik und Wissenschaftstheorie (Constructive Logic, Ethics and Philosophy of Science)*, Mannheim: Bibliographisches Institut, 2nd edn, 1975. (The ‘bible’ of Erlangen constructivism. The first part has a mature exposition of Lorenzen’s dialogical logic and the way it functions in the Erlangen programme mentioned in §3.)
- Walton, D.N. and Krabbe, E.C.W.** (1995) *Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning*, Albany, NY: State University of New York Press. (Pursues the integration of dialogical logic with other dialogue theory, especially Charles Hamblin’s formal dialectic.)

Dynamic logics

Dynamic logics have been designed by Pratt as formal systems for reasoning about computer programs. The main ingredients discussed are programs, operations on programs, states and properties of states. In particular one can formalize that every execution of a program p starting in state s terminates in a state with a given property. Thus correctness statements for programs can be dealt with. According to Segerberg (1980) programs might be viewed more generally as actions of some agent so that certain aspects of human action theory can also be formulated and studied in these systems.

1 Motivation

In Pratt's (1976) setting the meaning of a program p with respect to an abstract machine $\mathcal{M}$ is taken as the transition relation $\mathcal{M}_p$ between states of $\mathcal{M}$, so that $\mathcal{M}_p$ holds for states s and s' (symbolized by $s\mathcal{M}_p s'$) if and only if there is a machine run according to p starting in state s and terminating in state s' . The execution of programs can be non-deterministic - another run of $\mathcal{M}$ according to p might well lead from the same initial state to quite a different final state. Deterministic programs are therefore special programs where the transition relation is actually a (partial) function. From already given programs one can obtain other programs by applying operations like 'first do p_1 and then p_2 ' (*composition*), 'do p_1 or p_2 ' (*nondeterministic choice*) or 'apply p a finite number of times' (*iteration* or *looping*). Given a certain stock of basic propositions, i.e. properties of states, one can build new propositions by taking boolean combinations of already given propositions or asking modal questions like 'does every terminating run of a program p starting in state s necessarily end in a state with property Q ?'. Thus to every program p naturally corresponds a modal operator $[p]$... to be read: 'after (each terminating execution of) p it is the case that ...'. The proposition '*not*: $[p]$ false' expresses for instance at s that at least one run of p starting in s will terminate. Every proposition Q induces a test programme 'if Q proceed, otherwise skip' (*test*). The corresponding relation holds for s and s' if and only if $s = s'$ and s satisfies Q . Formalizing this situation leads immediately to propositional dynamic logic.

2 Propositional dynamic logic

The above remarks suggest the following formal language where formulas and program terms (for short: programs) are simultaneously built up from propositional and program variables with the aid of the usual propositional connectives, say $\neg$ (not) and $\rightarrow$ (implies), the p -necessity operator $[p]$ for programs p , and the program forming operations $;$ (composition), $\cup$ (choice), $*$ (iteration) and $?$ (test). The formation rules for formulas and programs are as follows:

- propositional variables are formulas and if φ and ψ are formulas and p is a program variable then $\neg\varphi$, $(\varphi \rightarrow \psi)$ and $[p]\varphi$ are formulas
- program variables are programs and if p and q are programs and φ is a formula then $(p;q)$, $(p \cup q)$, p^* and $\varphi?$ are programs.

Other propositional connectives and the p -possibility operator $\langle p \rangle$ are introduced by definition. For instance $\langle p \rangle$ stands for $\neg[p]\neg$. A Kripke-semantics is now easily established. A model $\mathcal{M} = (S, I)$ for this language consists of a nonempty set S (the set of states or 'possible worlds' of $\mathcal{M}$) and a function I which assigns to every propositional variable a property in S , i.e. a subset of S , and to every program variable a two-place relation in S , an accessibility relation. Then the truth of formulas φ in $\mathcal{M}$ at state s (symbolically: $\mathcal{M} \models_s \varphi$) and the interpretation of programs p (symbolically: $\mathcal{M}_p$) can be simultaneously defined. The interesting cases are:

- $\mathcal{M} \models_s [p]\varphi$ iff for every s' , $s\mathcal{M}_p s'$ implies $\mathcal{M} \models_{s'} \varphi$
- $\mathcal{M}_{(p;q)} = \mathcal{M}_p \circ \mathcal{M}_q$, $\mathcal{M}_{(p \cup q)} = \mathcal{M}_p \cup \mathcal{M}_q$, $\mathcal{M}_{p^*} = (\mathcal{M}_p)^*$
- $\mathcal{M}_{\varphi?} = \{(s, s) \mid \mathcal{M} \models_s \varphi\}$

where $R_1 \circ R_2$ denotes the product of R_1 and R_2 , i.e. $sR_1 \circ R_2 s'$ holds iff there is an s_1 such that $sR_1 s_1$ and $s_1 R_2 s'$, $R_1 \cup R_2$ denotes the set theoretical union of R_1 and R_2 , and R^* denotes the reflexive and transitive closure of R , i.e. $sR^* s'$ holds iff either $s = s'$ or for some natural number n there are $s_1, \dots, s_n$ such that $sR s_1$ and $\dots$ and $s_n R s'$ hold. A lot of constructs known from ordinary programming languages can now be formalized,

for instance ‘if φ then p else q ’ is $((\varphi?; p) \cup (\neg\varphi?; q))$ or ‘while φ do p ’ is $((\varphi?; p)^*; \neg\varphi?)$. The set of valid formulas, i.e. formulas which are true in every model at all states, can be generated (in Hilbert-style) from certain axioms by *modus ponens* and the necessitation rule. The set of all tautologies together with the formulas given by the schemas below could be chosen (after **Seegerberg 1980**) as axioms:

- $[p](\varphi \rightarrow \psi) \rightarrow ([p]\varphi \rightarrow [p]\psi)$
- $[(p; q)]\varphi \leftrightarrow [p][q]\varphi$
- $[(p \cup q)]\varphi \leftrightarrow ([p]\varphi \wedge [q]\varphi)$
- $[\varphi?]\psi \leftrightarrow (\varphi \rightarrow \psi)$
- $[p*]\varphi \rightarrow (\varphi \wedge [p][p*]\varphi)$
- $[p*](\varphi \rightarrow [p]\varphi) \rightarrow (\varphi \rightarrow [p*]\varphi)$

Correctness of the system is easily verified and completeness has been proved by Gabby, Parikh and Seegerberg (see **Goldblatt 1987**). There are also correct and complete Gentzen-type systems. The question of general validity of a formula turns out to be decidable (**Fischer and Ladner 1979**).

3 Dynamic predicate logic

Viewing states as the content of certain storage units $x_0, x_1, x_2, \dots$ they can be identified with functions assigning to each x_n its contents $s(x_n)$. Thus $s(x_n)$ denotes the object (for instance a string over an alphabet) that occupies in state s the n -th storage unit. The set of objects that are allowed to appear in some storage unit constitutes a one- (or many-) sorted domain in which certain basic first- (or higher-) order properties and functions are fixed. Using these initial data a first- (or higher-) order possibly many-sorted modal language, known as dynamic predicate logic, can be used to describe transition phenomena initiated by programs in state sets. Simple assignments of the form $x := t$ (to be read: ‘set x equal to t ’) are often taken as the only basic programs from which all others are derived by program forming operations. A run of such a simple assignment $x := t$ starting in state s has the effect that only the content of the storage unit x is ‘updated’ by the value of t and every other unit is left untouched. The resulting state s' will be denoted by $s(x/t)$. Simple assignment programs are obviously deterministic and universally terminating. Fixing a first order language with individual variables $x_0, x_1, x_2, \dots$ and certain relation and function symbols, a model $\mathcal{M} = (D, I)$ for such a language consists of a nonempty domain D and an interpretation function I which associates with every relation and function symbol of the language an appropriate relation or function in D . Every formula φ then defines a property in S , the set of all assignments to the variables - that is the set of all functions from $\{x_0, x_1, x_2, \dots\}$ to D . An element s in S has the φ -property iff φ is true in $\mathcal{M}$ under the assignment s , symbolically $\mathcal{M} \models_s \varphi$. The essential new case in defining truth is:

- $\mathcal{M} \models_s [x := t]\varphi$ iff $\mathcal{M} \models_{s(x/t)} \varphi$

Therefore $[x := t]\varphi$ is (under certain syntactical conditions) equivalent to $\varphi(x/t)$ which arises from φ by substituting t for every free occurrence of x in φ . Using the iteration operation it is possible to express things which in general cannot be said by any fixed number of applications of the substitution operation. For instance $[(x_1 := f(x_1))^*]Q(x_1)$ expresses in $\mathcal{M}$ at s that all of

$$s(x_1), f^{\mathcal{M}}(f^{\mathcal{M}}(s(x_1))), \dots$$

have the property corresponding to Q in $\mathcal{M}$. In such a language the natural numbers can now be characterized up to isomorphism by a formula. The main part of this formula says that for every y the program

$$(x := 0; \text{ while } x \neq y \text{ do } x := x + 1)$$

terminates (at least once). This means that every element can be reached in finitely many steps from 0 by adding 1 each time. Here iteration is used essentially. From the characterizability of the natural numbers it follows in particular that the set of valid formulas is not recursively enumerable so that no usual proof system can be correct and complete. However, by adding an infinitary rule (a sort of ω -rule) correctness and completeness can be re-established.

Dynamic predicate logic can be viewed as a sublogic of an extended predicate logic in which countable

conjunctions and disjunctions are allowed. That is because a formula $[p*]\varphi$ is obviously equivalent of the disjunction over the set of formulas.

$\varphi, [p]\varphi, [(p;p)]\varphi, [((p;p);p)]\varphi, [(((p;p);p);p)]\varphi, \dots$

Many variations of the core formalism are possible; for instance tests are often restricted to Boolean combinations of atomic formulas. This seems to be more realistic since tests of the form $(\forall x\varphi)?$ might involve a search through an infinite domain. Also, random assignments (written as $x:=?$) are sometimes introduced. In a model $\mathcal{M} = (D, I)$ such a basic random assignment denotes the relation which holds between two assignments s and s' if and only if $s' = s(x/d)$ for some $d \in D$. The formula $\forall x\varphi$ can then be expressed by $[x:=?]\varphi$.

See also: [Logical and mathematical terms, glossary of](#)

ULF FRIEDRICHS DORF

References and further reading

- Fischer, M.J. and Ladner, R.E.** (1979) 'Propositional dynamic logic of regular programs', *Journal of Computer and System Sciences* 18, 194-211. (Presents the first proof of decidability for propositional dynamic logic.)
- Goldblatt, R.** (1987) 'Logics of Time and Computation', *CSLI Lecture Notes Number 7*, Stanford. (Good general reference for modal, tense and dynamic logic.)
- Harel, D.** (1984) 'Dynamic logic', in D. Gabbay and F. Guenther (eds) *Handbook of Philosophical Logic* vol. 2, Dordrecht: Reidel, 497-60. (Extensive handbook article presenting many results and some proofs. Contains valuable bibliographic information.)
- Kozen, D. and Tiuryn, J.** (1990) 'Logics of programs', in J. v. Leeuwen (ed.) *Handbook of Theoretical Computer Science* vol. B, Amsterdam: Elsevier, 789-840. (Another, perhaps more accessible, handbook of articles.)
- Pratt, V.R.** (1976) 'Semantical considerations on Floyd-Hoare logic', *Proceedings of 17th IEEE Symposium on Foundations of Computer Science*, 109-21. (Work initiating the subject.)
- Segerberg, K.** (1980) 'Applying modal logic', *Studia Logica* vol. 39, 275-95. (A good introduction to dynamic logic.)

Epistemic logic

Modern treatment of epistemic logic began in the 1950s when some philosophers noticed (as scholastics had done before them) certain regularities in the logical behaviour of the concept of knowledge (for example, that knowing a conjunction is equivalent to knowing all its conjuncts) and began to systematize them. Initially these regularities were presented in the form of an axiomatic-deductive system, as in other branches of logic. Later, questions began to be asked concerning the model theory on which such an 'epistemic logic' is based. Still later, the concrete interpretation of this model theory has become an issue. In this way, gradually a bridge has begun to be forged from purely logical questions to such central epistemological questions as those concerning the objects of knowledge, different kinds (or even senses) of knowledge (and their interrelations), the intensional character of knowledge, the de dicto versus de re distinction, and so on.

1 Propositional epistemic logic

Epistemic logic began as a study of the logical behaviour of the expression 'knows that' (see **Propositional attitude statements**). One of the main aims of this study is to be able to analyse other constructions in terms of 'knows' by means of '*b* knows that'. This basic notion will be expressed in the notation used here by ' K_b '. (This symbolization is slightly misleading in that, in a formula of the form $K_b S$, the term '*b*' for the agent (knower) is intended to be outside the scope of ' K ', not inside, as the notation might suggest.) When the knower is the same throughout a sentence (or an argument), the '*b*' is often omitted.

The logic of ' K ' ('knows that') is based on certain simple, pragmatic and epistemological insights. First, what is the use of the notion of knowledge? In what way are you better off when you come to know that *S*? The basic answer is clear. When you know that *S*, you can legitimately omit from consideration all possibilities under which it is not the case that *S*; in other words, you can restrict your attention to the situations in which it is true that *S*. (What is meant by 'legitimately' here is a major philosophical problem. It is essentially tantamount to the well-known problem of defining knowledge (see **Knowledge, concept of**). For the purposes of epistemic logic, however, we do not have to solve this problem. The structural characterization just given suffices for our purposes, no matter how legitimacy is defined.)

This characterization amounts to making a distinction between two kinds of alternatives (situations; possible worlds) among all the relevant ones. Some possibilities are excluded by what *b* knows. They are the ones that can legitimately be disregarded. Others are not so excluded. They are the ones *b* will have to heed. It should be emphasized that the specification of such a set of admitted situations can be made independently of the psychological, cognitive or sociological peculiarities of the knower in question. Hence epistemic logic does not necessarily involve any subjective or even pragmatic element.

Thus we can see what kinds of structures are to be used in the model theory of epistemic logic (see **Model theory**). They will be called 'model structures'. They are classes of models of the underlying language minus *K*. Let us consider one of them, say, w_0 , in which a potential knower, *b*, exists. Each such w_0 is associated with a number of other models ('possible worlds') w_1 which are not excluded by what *b* knows in w_0 . They will be called the epistemic *b*-alternatives to w_0 . Thus the basic truth-condition of epistemic logic can be formulated as follows:

$K_b S$ is true in w_0 iff *S* is true in every epistemic *b*-alternative w_1 to w_0 .

This definition is of course relative to a model structure. If the *b*-alternativeness relation is called R_b and Ω is the set of all possible worlds, the truth-condition can be expressed as follows:

$K_b S$ is true in $w_0 \in \Omega$ iff *S* is true in each $w_1 \in \Omega$ such that $w_0 R_b w_1$.

We can also introduce a notion of epistemic possibility, *P*. Then $P_b S$ means, roughly, 'It is possible, for all that *b* knows, that *S*'. Obviously the following pairs of sentences are logically equivalent: $\sim K_b S$ and $P_b \sim S$; $\sim P_b S$ and $K_b \sim S$. The truth-condition for epistemic possibility is the obvious one:

$P_b S$ is true in $w_0 \in \Omega$ iff *S* is true in some $w_1 \in \Omega$ such that $w_0 R_b w_1$.

Normally it is required that each $w_0 \in \Omega$ be an alternative to itself (for each *b* in w_0). This amounts to requiring

that b can know (in w_0) that S only if it is true (in w_0) that S . Combined with the usual truth-conditions for propositional connectives, we can in this way obtain a semantic basis, complete with a truth-definition, for propositional epistemic logic.

The question of whether the alternativeness relation is transitive or not is equivalent to the question of the logical truth of sentences of the form:

$$(1) (\forall x)(K_x S \supset K_x K_x S).$$

Unfortunately, in ordinary discourse, saying that one ‘knows that one knows’ has a number of different meanings, most of which have little to do with (1). Hence the question of whether ‘knowing implies that one knows that one knows’ needs a separate examination which falls only partially within the purview of epistemic logic.

On the basis of these semantic ideas, an epistemic propositional logic is easily developed. It admits of a complete Gentzen-style axiomatization. In such an axiomatization, the rules for ‘ $\&$ ’ and ‘ $\vee$ ’ are the usual cut-free ones, while no transfer of formulas between the left-hand side and the right-hand side is allowed. (For simplicity, it is assumed here and in the following that the formulas considered are in negation normal form, that is, that all negation signs in them are prefixed to atomic formulas.) The epistemic rules can then be formulated as follows:

$$\frac{\Gamma, K(S_1 \& S_2) \rightarrow \Delta}{\Gamma, K S_1, K S_2 \rightarrow \Delta}$$

$$\frac{\Gamma \rightarrow \Delta, P(S_1 \vee S_2)}{\Gamma \rightarrow \Delta, P S_1, P S_2}$$

$$\frac{\Gamma \rightarrow \Delta, P S_1, P S_2}{S_1, S_2 \rightarrow S_3}$$

$$\frac{\Gamma, K S_1, P S_2 \rightarrow \Delta, P S_3}{S_1 \rightarrow S_2, S_3}$$

$$\frac{\Gamma, K S_1 \rightarrow \Delta, K S_2, P S_3}{\Gamma, K S_1 \rightarrow \Delta, K S_2, P S_3}$$

We also need, of course, suitable structural rules and axioms. The following can serve as such axioms.

$$\frac{S_1, \sim S_1 \rightarrow}{\rightarrow S_1, \sim S_1}$$

$$S \rightarrow S$$

A game-theoretic truth-definition is better suited for the purposes of epistemic logic than an ordinary (Tarski-type) one (see **Semantics, game-theoretic**). In such a definition two players are involved: the proponent and the opponent. At each stage of the game, the two players are considering a sentence and a possible world. The game starts from the world w_0 in which a sentence S is to be evaluated. Such a game $G(S; w_0)$ is defined by the following rules:

- (G. $\sim$) $G(\sim S; w_0)$ is like $G(S; w_0)$ but with the roles of the two players reversed. (That is, the proponent of $G(\sim S; w_0)$ is the opponent in $G(S; w_0)$, and vice versa.)
- (G. $\&$) $G((S_1 \& S_2); w_0)$ begins with a choice by the opponent of S_i ($i = 1$ or 2) and the rest of the game is as in $G(S_i; w_0)$.
- (G. $\vee$) $G((S_1 \vee S_2); w_0)$ likewise, except that the choice of S_1 or S_2 is made by the proponent.
- (G.K) $G(K_b S; w_0)$ begins with a choice by the opponent of an epistemic b -alternative w_1 to w_0 . The rest of the game is as in $G(S; w_1)$.
- (G.P) $G(P_b S; w_0)$ likewise, except that the choice is made by the proponent.
- (G.atom) Any play of a semantic game will come to an end in a finite number of moves with an atomic sentence or identity A and a model (world) w . Since the underlying language has been interpreted in w , A is either true or false. If the former, the proponent of A has won and the opponent lost. If the latter, vice versa.

The game-theoretic definition of truth and falsity says that S is true (false) in w_0 iff there exists a winning strategy for the proponent (opponent) in $G(S; w_0)$.

2 The semantics of quantified epistemic logic

When we move to *quantified* epistemic logic, we need individuals to serve as values of bound variables. Since we

are dealing with a multiplicity of models (worlds), individuals must in principle be identifiable in more than one of them. ('No entity without identity', to quote Quine.) Hence an individual is, model-theoretically speaking, a 'world line' connecting the roles or embodiments of one and the same individual in several worlds. Less metaphysically expressed, given a model system Ω , an individual is logically speaking a function (an 'individuating function') from models $w \in \Omega$ to the members of the domain $do(w)$ of w (see [Intensional logics §2](#)). (Since we want to keep our discussion as general as possible, we must allow individuating functions to be partial. In other words, an individual might not be identifiable in all possible worlds.)

A singular term likewise has as its value (bearer), in a given model w , a member of $do(w)$. However, these might not be the same individual in two different models. In other words, the values of a singular term c in two worlds might not be connected by a world line. The bearers of a singular term c in different models are said to be the values of the 'meaning function' associated with c . Individuating functions are a special case of meaning functions. Obviously, a meaning function is, like an individuating function, a (possibly partial) function from models $w \in \Omega$ to their respective domains $do(w)$. (In the interests of generality, a singular term must be allowed to be empty in some models.)

It cannot be assumed in epistemic logic that individuating functions are expressed in ordinary language by singular terms of any particular linguistic category, such as proper names, in the sense that the individuating function would coincide with the meaning function of such a 'rigid designator'. This impossibility is shown by the fact that one can meaningfully ask, using a proper name, for example, 'Do you know who John is?' or 'Who is John?' The idea that there are in natural language 'rigid designators' that express individuating functions hence does not work in epistemic contexts (see [Reference §3](#)).

On the basis of these considerations, we can now formulate the additional game rules needed in epistemic first-order logic:

(G. $\forall$) $G((\forall x)S(x); w_0)$ begins with the choice by the opponent of an individual from the domain of w_0 . If it is stipulated that a new singular term c has the same meaning function as the identifying function of the chosen individual, the rest of the game is as in $G(S(c); w_0)$.

Even though it seems to restrict our treatment somewhat, we can also require that the chosen individual exists in w_0 , that is, that its identifying function is defined for the argument w_0 .

(G. $\exists$) $G((\exists x)S(x); w_0)$ likewise, except that the proponent makes the choice.

In the epistemic first-order logic so far formulated, many interesting logical relationships can be expressed. However, a fully satisfactory logic of the concept of knowledge can be obtained only by extending our language in one important respect. For one thing, the interrelationship between our formal language and knowledge statements in ordinary language is made much more perspicuous by this extension.

The new element in our epistemic language is a notation for informational independence, slightly extended. Moves are sometimes made by a player in ignorance of what has happened in certain specified earlier moves. Such moves are said to be independent of the earlier ones. This idea applies naturally to our semantic games. In them an application of the rules (G. $\&$), (G. $\vee$), (G. $\forall$) or (G. $\exists$) can be informationally independent of an earlier application of one of the rules (G. K) or (G. P). Actually, in much of epistemic logic, we need to consider only the possible independence of (G. $\vee$) and (G. $\exists$) of (G. K).

Independence will be indicated by a slash notation, for example, in ($\vee/K$) and ($\exists x/K$). However, the possibility of moving from one model to another necessitates a specification as to what the model w is in relation to which a move connected with, say, ($\exists x/K$) is made. Here the obvious stipulation is that it must be made in relation to the model (world) which the rule application connected with K moved us away from. In brief, the slash indicates that the moves connected with K and with the existential quantifier $\exists x$ are made in the reverse order.

These stipulations imply that there is a difference in meaning, for example, between sentences of the following forms: $K(\forall x)S(x)$ and $(\forall x)KS(x)$. In the former, the choice of the values of x is between the individuals existing in an epistemic alternative chosen by the opponent. In the latter, the choice is between individuals existing in the world in which the sentence is evaluated.

It is our reliance on the notion of informational independence that forces us to use game-theoretic semantics. The reason is that the logic of informational independence cannot be handled by means of the usual Tarski-type truth-definitions.

By means of the slash notation, most of the usual types of knowledge, and most of the logical laws holding for these different types of knowledge, can be formulated. The value of a quantifier such as ‘ $(\exists x/K_b)$ ’ is chosen independently of the epistemic alternative selected by the opponent in applying the rule (G.K). Hence the value of x must be the same in all the situations compatible with what b knows, that is, must be known to b . Yet this value may depend on that of a universal quantifier which depends on K , wherefore no linear ordering can capture the dependence relations in question. Thus we have, for instance:

- (2) ‘ b knows whether S ’ $= K_b(S(\vee/K_b)\sim S)$
- (3) ‘ b knows whether S_1 or S_2 ’ $= K_b(S_1(\vee/K_b)S_2)$
- (4) ‘ b knows who satisfies the condition $S(x)$ ’ $= K_b(\exists x/K_b)S(x)$
- (5) ‘ b knows of everybody whether they satisfy the condition $S(x)$ ’ $= K_b(\forall x)(S(x)(\vee/K_b)\sim S(x))$
- (6) ‘ b knows to whom each person bears the relation $S(x, y)$ ’ $= K_b(\forall x)(\exists y/K_b)S(x, y)$.

Even more complex types of knowledge can be expressed in our notation. In these examples, it is assumed that the range of quantifiers consists of persons.

Of (2)-(6), the first three have equivalents in a slash-free notation, namely:

- (7) $K_b S \vee K_b \sim S$
- (8) $K_b S_1 \vee K_b S_2$
- (9) $(\exists x)K_b S(x)$.

However, (5) and (6) do not have slash-free equivalents. For instance, (6) is not equivalent to:

- (10) $(\forall x)(\exists y)K_b S(x, y)$.

This is seen from the following second-order equivalents to (6) and (10), respectively:

- (11) $(\exists f)K_b(\forall x)S(x, f(x))$
- (12) $(\exists f)(\forall x)K_b S(x, f(x))$.

This reduction of a large number of other constructions with ‘knows’ to the ‘knows that’ construction is evidence for the propositional character of knowledge.

3 Rules for quantified epistemic logic

In terms of the slash notation, deductive rules for quantified epistemic logic can be formulated.

First, the vaunted intensionality of knowledge can be understood and accounted for without postulating any special intensional entities to serve as ‘objects of knowledge’ (see **Intensional entities**). Intensionality is a matter of how meaning functions and identifying functions behave. It means that an identity which holds in one situation (model) need not hold in another one. In other words, we cannot have in our epistemic logic an unrestricted rule of the following form:

$$(13) \frac{\Gamma, S(a), a = b, S(b) \rightarrow \Delta}{\Gamma, S(a), a = b \rightarrow \Delta}$$

We can have such a rule only when S is required not to contain any intensional operators.

Second, rules for existential generalization and universal instantiation have to be changed. For instance, we cannot have the rule:

$$(14) \frac{\Gamma, (\forall x)S(x), S(a) \rightarrow \Delta}{\Gamma, (\forall x)S(x) \rightarrow \Delta}$$

for this inference presupposes that a is the same individual in all the relevant models. If x occurs in $S(x)$ within

the scope of only one epistemic operator, say, K (while $S(x)$ is in negation normal form), this rule can be restored by means of an additional premise.

$$(15) \frac{\Gamma, (\forall x)S(x), K(\exists y/K)(a = y), S(a) \rightarrow \Delta}{\Gamma, (\forall x)S(x), K(\exists y/K)(a = y) \rightarrow \Delta}$$

This modification of the usual rules is eminently natural both ‘intuitively’ and model-theoretically. If I know that a did something, it does not follow that I know who did it, unless I know who a is. At the same time, the need of the extra clause shows that in a sense the objects of knowledge are not world-bound individuals, but individuals that can be identified in several possible situations. If this is what is meant by saying that knowledge is intentional, the saying is true.

Third, rules for existential instantiation and universal generalization also have to be changed. They no longer do their whole job. We cannot have merely:

$$(16) \frac{\Gamma, S(a) \rightarrow \Delta}{\Gamma, (\exists x)S(x) \rightarrow \Delta}$$

where a does not occur anywhere in $S(x)$, Γ or Δ . Instead, we need a stronger rule of the form:

$$(17) \frac{\Gamma, S_0(S_1(f(y_1, y_2, \dots))) \rightarrow \Delta}{\Gamma, S_0((\exists x)S_1(x)) \rightarrow \Delta}$$

where S_0 is in negation normal form and $(\forall y_1), (\forall y_2), \dots$ are all the universal quantifiers within the scope of which $(\exists x)S_1(x)$ occurs in $S_0 = S_0((\exists x)S_1(x))$, and where f is a function symbol which does not occur anywhere in S_0 , Γ or Δ . Clearly (16) can be thought of as a special case of (17). Parallel sequent rules have to be modified analogously.

Fourth, negation behaves in quantified epistemic logic in an unwonted way. For instance, the law of excluded middle no longer holds. This is almost predictable on the basis of our definitions of truth and falsity in §1 above. They imply that the law of excluded middle is a determinacy assumption, that is, that it says that one or the other player always has a winning strategy in one of our epistemic games. From game theory it is known that such determinacy assumptions are very strong, and hence likely to fail.

In general, quantified epistemic logic thus exhibits several phenomena that have no counterpart in ordinary first-order logic. In spite of the failure of the law of excluded middle, one can formulate a complete disproof procedure for epistemic logic. It can also be shown that many standard metalogical results, such as compactness, and the separation and Löwenheim-Skolem theorems, can be extended to epistemic logic.

4 Wider perspectives

The epistemic logic outlined here can be extended and applied in several different directions.

First, epistemic logics have proved relevant to such branches of computer science as artificial intelligence and database theory. A growing number of results have been reached by scientists working in this area.

Second, an interpretational rule (game rule) must be associated also with nonlogical constants of different kinds, for instance, with singular terms such as names. Applications of such rules can be independent of epistemic rules, which create distinctions between sentences such as the following:

$$(18) K_b S(d)$$

$$(19) K_b S((d/K))$$

This distinction is nothing but the so-called *de re-de dicto* distinction (see *De re/de dicto*). This famous distinction turns out to be another independence phenomenon and not an irreducible contrast between two kinds of knowledge nor two kinds of reference.

Third, epistemic logic is the cornerstone of the logic of knowledge-seeking developed by Jaakko Hintikka and his associates and called by them the interrogative model of inquiry (*Hintikka et al. 1997*). In fact, we obtain the rules

for such an inquiry by reversing Gentzen-type rules for epistemic logic and adding to them the following additional rule:

$$(R.Q) \frac{\Gamma, S_0 \rightarrow \Delta}{\Gamma, S_0, S_1 \rightarrow \Delta}$$

where S_0 is the presupposition of any one question and S_1 the desideratum of the same question. In such interrogative inquiry, it must also be specified which questions are answerable. The applicability of (R.Q) then presupposes the answerability of the question.

Fourth, in epistemic contexts quantifiers have been shown to depend on a method of identifying individuals as members of different possible situations. Now it can be seen that individuals can be identified in two different ways. Then we have two sets of identifying functions and, corresponding to them, two pairs of quantifiers. The rules governing each of them separately are the same as have been explained. In one kind of identification, the knower uses their direct cognitive relations to objects (for example, their visual space) as a framework of identification. Such identification has been called ‘perspectival’. In ordinary language, perspectival identification shows up in the form of the direct-object construction with ‘knows’, for example, ‘Jack knows Jill’, as distinguished from ‘Jack knows who Jill is’. In the other, an impersonal (public) framework of identification is relied on. It can be shown that this duality is indeed realized in our actual conceptual practice with epistemic concepts. This results in an extremely important application of epistemic logic. In this application, questions concerning the interaction of the two kinds of quantifiers belong to the province of epistemology and cognitive science rather than epistemic logic.

This distinction between two kinds of identification is approximately instantiated by Bertrand Russell’s distinction between knowledge by description and knowledge by acquaintance (see [Knowledge by acquaintance and description](#)), by neuroscientists’ distinction between the what-system and where-system in cognition, by Endel Tulving’s distinction between semantic and episodic memory, and so on. It is important to realize that these distinctions are not at bottom distinctions between two kinds of knowledge (or memory-knowledge), or between two kinds of objects within some one world, but between two kinds of principles of identification of individuals from one world to another. It is nevertheless tempting and even natural to speak of ‘two kinds of objects’ here. For instance, so-called phenomenological objects (including Russell’s ‘objects of acquaintance’) are in effect individuals identified in a certain way.

Fifth, in natural language we are dealing with more than just one domain of individuals. Rather, we have at the very least a different domain corresponding to each different ‘wh- question’. This distinction between different ranges of quantifiers (for that is what wh- words in effect express) is closely related to Aristotle’s distinction between different categories.

Finally, a model-theoretic treatment of knowledge seems to lead to the paradox of logical omniscience, that is, to the consequence that everyone knows all the logical consequences of what they know. One natural explanation is that in their actual thinking humans articulate the possible worlds they are tacitly considering only to a certain depth. This notion of a world articulated only to a given depth can be captured by means of Rantala’s (1975) notion of urn models, which hence offers a natural way out of the paradox.

See also: [Logic and mathematical terms, glossary of](#)

JAAKKO HINTIKKA
ILPO HALONEN

References and further reading

- Boh, I.** (1993) *Epistemic Logic in the Later Middle Ages*, London: Routledge. (The prehistory of epistemic logic.)
- Fagin, R., Halpern, J.Y., Moses, Y. and Vardi, Moshe Y.** (1985) *Reasoning about Knowledge*, Cambridge, MA: MIT Press. (Synthesis of computer scientists’ work on epistemic logic.)
- Hintikka, J.** (1962) *Knowledge and Belief*, Ithaca, NY: Cornell University Press. (First fully fledged treatment of the logic of knowledge.)
- Hintikka, J.** (1992) ‘Independence-Friendly Logic as a Medium of Information Representation and Reasoning About Knowledge’, in S. Ohsuga *et al.* (eds) *Information, Modelling and Knowledge Bases*, Amsterdam: IOS Press, vol. 3, 258- 65. (Develops the themes of §§2-3 of this entry.)

- Hintikka, J.** (1996) 'Knowledge Acknowledged: Knowledge of Propositions vs. Knowledge of Objects', *Philosophy and Phenomenological Research* 56: 251-75.(Knowledge of truths and knowledge of objects distinguished from each other.)
- Hintikka, J., Halonen, I. and Mutanen, A.** (1997) 'Interrogative logic as a general theory of reasoning', in R. Johnson and J. Woods (eds) *Handbook of Applied Logic*, Dordrecht: Kluwer.(Survey of the uses of epistemic logic in knowledge acquisition by questioning.)
- Hintikka, J. and Hintikka, M.B.** (1989) *The Logic of Epistemology and the Epistemology of Logic*, Dordrecht: Kluwer.(Essay 2 is an alternative survey of epistemic logic; essay 4 applies epistemic logic to the problem of the objects of knowledge; essay 5 treats the problem of logical omniscience; and essay 8 explores the connections between different modes of identification and their manifestations in cognitive science.)
- Hintikka, J. and Sandu, G.** (1989) 'Informational Independence as a Semantical Phenomenon', in J.E. Fenstad, I.T. Frolov and R. Hilpinen (eds) *Logic, Methodology and Philosophy of Science*, Amsterdam: North Holland, vol. 8, 571-89.(Introduces the idea of informational independence into logical semantics.)
- Lenzen, W.** (1978) *Recent Work in Epistemic Logic*, Acta Philosophica Fennica, vol. 30, number 1, Helsinki: Societas Philosophica Fennica.(Survey of the literature up to 1978.)
- Rantala, V.** (1975) 'Urn Models', *Journal of Philosophical Logic* 4: 455-74.(Urn models defined.)
- Wright, G.H. von** (1951) *An Essay in Modal Logic*, Amsterdam: North Holland.(Von Wright began the first systematic treatment of epistemic logic.)

Existence

Philosophical problems concerning existence fall under two main headings: 'What is existence?' and 'What things exist?'. Although these questions cannot be entirely separated, this entry will concentrate on the first.

The question 'What is existence?' has produced a surprising variety of answers. Some hold that existence is a property that every individual has, others that it is a characteristic that some individuals have but other (for example, imaginary) individuals lack, while proponents of the thesis that 'existence is not a predicate' hold that existence is not a property or characteristic of individuals at all.

Other philosophical issues concerning existence include: disputes about whether there are abstract objects (for example, numbers, universals) as well as concrete ones, immaterial souls as well as bodies, possible objects as well as actual ones, and so on; and questions about which entities (if any) are the fundamental constituents of reality.

1 Objects and existence

To the question 'What is existence?' it is tempting to respond: 'A property that belongs to everything'. However, this answer appears to have paradoxical consequences.

An ancient puzzle, traceable to Parmenides, concerns denials of existence. If existence is a property of everything, one cannot talk about anything that does not exist. But if one cannot talk about what does not exist, how can one truly deny the existence of anything? Yet it is obvious that there are true denials of existence (true negative existential statements): Santa Claus does not exist, nor do Henry VIII's seventh wife or talking tigers.

A second puzzle concerns assertions of existence (positive existential statements). If everything that can be talked about exists, how can a true assertion that something exists be other than trivial? Yet it is obvious that there can be true but non-trivial assertions of existence: 'Egg-laying mammals exist', for example, is not trivial; and 'Homer existed' would be informative if true.

One response to these puzzles consists in holding that as well as objects that exist, there are objects that do not exist. Theories of this type are called 'Meinongian' (see [Meinong 1904](#)), after the most famous exponent of such a theory (see [Meinong, A.](#)). According to a Meinongian theory, the possibility of true negative existential statements is explained by the fact that Santa Claus, the seventh wife of Henry VIII and so on are objects which can be talked about and can possess various properties (such as driving reindeer and being married), but which lack existence. As for the second puzzle: if some objects are nonexistent, it can clearly be informative to be told, of an object, that it exists.

Such a theory may appear to offer an attractive account of discourse about fiction; one that allows us to take at face value the fact that both 'Pegasus has wings' and 'Pegasus does not exist' appear to be true. However, there are features of this Meinongian theory that many find unacceptable. For example, if we are to use the theory to explain all true negative existential statements, we must follow Meinong in admitting that there are impossible objects such as the round square (since 'The object that is both round and square does not exist' is evidently true). Also, the Meinongian theory implies that Pegasus, although nonexistent, is a horse, yet it is debatable whether the property of being a horse can belong to something nonexistent. Even the fundamental tenet of the theory - that we can distinguish being an object from being an existent object - is open to question. If we admit that Pegasus is an object, a possible subject of discourse and a bearer of properties, does it make sense to add that Pegasus does not exist? If the answer is that existence requires location in space and time, although being an object does not, we may ask why we should accept this distinction.

2 The quantifier account of existence

An influential rival to the Meinongian account holds that the puzzles described in §1 above arise from the mistaken assumption that the term 'exist(s)' attributes a property to individuals. This solution employs an account of existence given by Frege and elaborated by Russell (see [Frege, G. §2](#); [Russell, B. §9](#)). The account is associated with the slogan 'existence is not a predicate', a dictum derived from the writings of Kant, who anticipated the Frege-Russell theory in claiming that the ontological argument for God's existence depends on the illegitimate

assumption that existence is a property (see [God, arguments for the existence of §§2-3](#)).

According to the Frege-Russell theory, the superficial grammatical form of the sentence ‘Egg-laying mammals exist’ (according to which it appears to attribute a property of existence to egg-laying mammals) is misleading. Its true logical form is revealed by reading it as ‘There are egg-laying mammals’, which can be analysed using logical symbolism as ‘ $(\exists x)x$ is an egg-laying mammal’, where ‘ $\exists x$ ’ is the existential quantifier (to be read as ‘There is some x such that...’). According to this ‘quantifier’ account of existence (or of the word ‘exist(s)’) the grammatical predicate ‘exist(s)’ disappears when the sentence is subjected to logical analysis, being replaced by the existential quantifier. As for ‘Talking tigers do not exist’, this is to be treated as equivalent to ‘There are no talking tigers’, which can be analysed as ‘It is not the case that: $(\exists x)x$ is a talking tiger’. Frege and Russell combined this quantifier account of existence with a claim to the effect that although ‘exist(s)’ is not a first-order predicate (a predicate of individuals), it is a higher-order predicate: a predicate of properties. For example, although in ‘Egg-laying mammals exist’ the word ‘exist’ does not ascribe the property of existence to egg-laying mammals, it does ascribe a property - that of being instantiated (‘having an instance’) - to the property of being an egg-laying mammal.

Using his celebrated theory of descriptions, Russell (1918) argued that the quantifier account can be extended to existential sentences involving definite descriptions and proper names, thus providing an ingenious general solution to the puzzles described above (see [Descriptions](#)). According to Russell, ‘Henry VIII’s seventh wife does not exist’ and ‘Santa Claus does not exist’ should be construed not as saying, of certain objects, that they lack existence, but rather as saying that certain properties (for example, being a seventh wife of Henry VIII or being a unique white-bearded, reindeer-driving, gift-delivering inhabitant of the North Pole) are not instantiated; that nothing has these properties. Similarly, ‘God does not exist’ is to be interpreted as the claim that a certain property - perhaps that of being a uniquely powerful, wise and benevolent creator of the universe - is not possessed by anything. Finally, if, following Russell, we treat the proper name ‘Homer’ as equivalent to the description ‘the author of the *Iliad* and the *Odyssey*’, we can construe the assertion that Homer existed as the claim that one person wrote both of these poems, an item of information that would be far from trivial.

The quantifier account is widely accepted as an analysis of general existential statements. However, many philosophers have been reluctant to follow Russell in holding that all meaningful existential statements are to be analysed in terms of the existential quantifier. (A notable exception is Quine (1953).) One difficulty is that there appear to be meaningful statements asserting the existence of particular things that resist the quantifier treatment. Even if we accept Russell’s (controversial) analysis of definite descriptions, and accept his descriptive analysis of proper names such as ‘Santa Claus’ and ‘God’, there are powerful reasons for thinking that most proper names are not equivalent to descriptions (see [Proper names §§1-3](#)). Hence it is doubtful whether ‘Bertrand Russell existed’ is equivalent to any claim that some set of properties is or was instantiated. Even more recalcitrant are existential statements involving demonstratives, such as ‘This exists’ or ‘That elephant exists’. Unless we say, implausibly, that all such sentences are false or meaningless, we appear to be committed to the conclusion that there are, after all, some true sentences in which ‘exist(s)’ ascribes a property of existence to particular objects. Further, it has been argued that some statements saying that things have existed, will exist, or might not have existed cannot be given a satisfactory analysis without invoking a property of existence.

However, if existence is a genuine property, an account must be given of the relation between this property and the existential quantifier, since it is undeniable that some general existential sentences are closely related to, if not equivalent to, existential quantifications. Moreover, if ‘exist(s)’ can express a property of individuals, does it play this role in any of the negative and positive existential statements discussed in §1? If so, how - assuming that the Meinongian account is rejected - are the puzzles to be solved? And what, if anything, can be said about the alleged property of existence beyond the claim that it belongs to everything? There is at present no consensus about these issues.

See also: [Being](#); [Fiction, semantics of](#); [Free logics](#); [Ontology](#); [Ontological commitment](#); [Logical and mathematical terms, glossary of](#)

PENELOPE MACKIE

References and further reading

- Barnes, J.** (1972) *The Ontological Argument*, London: Macmillan, ch. 3. (A succinct critical discussion of the thesis that existence is not a predicate, including many references to other works.)
- Evans, G.** (1982) *The Varieties of Reference*, Oxford: Clarendon Press, ch. 10. (Defends the thesis that existence is a property of everything, offering a sophisticated solution to some of the problems mentioned in §2.)
- Meinong, A.** (1904) 'Über Gegenstandstheorie' (The Theory of Objects), in R.M. Chisholm (ed.) *Realism and the Background of Phenomenology*, New York: Free Press, 1960, 76-117. (A difficult work. Parsons' *Nonexistent Objects* is an easier introduction to Meinongian views.)
- Parsons, T.** (1980) *Nonexistent Objects*, New Haven, CT: Yale University Press. (A modern and thorough development of Meinong's views. Quite technical in places.)
- Quine, W.V.** (1953) 'On What There Is', in *From a Logical Point of View: Nine Logico-Philosophical Essays*, Cambridge, MA: Harvard University Press, 2nd edn, 1980. (A classic piece of advocacy for the quantifier account of existence, introducing Quine's famous 'criterion of ontological commitment'.)
- Russell, B.A.W.** (1918) 'The Philosophy of Logical Atomism', Lectures 5 and 6, in *Logic and Knowledge: Essays 1901-1950*, ed. R.C. Marsh, London: Routledge, 1992. (An informal presentation by Russell of his version of the quantifier account.)
- Williams, C.J.F.** (1981) *What is Existence?*, Oxford: Clarendon Press. (A wide-ranging discussion, including a defence of a version of the quantifier account against numerous objections.)

Fallacies

Fallacies are common types of arguments that have a strong tendency to go badly wrong, or to be used as deceptive tricks when two parties reason together. In some instances they are simply careless errors in thinking, while in other cases they are techniques of argumentation used by one arguer more or less deliberately to try to fool another into accepting a false conclusion. Fallacies have been described in logic textbooks since the time of Aristotle; their study, long neglected, has begun to be revived in recent years, as its practical importance in natural language reasoning has become apparent.

Many fallacies have been recognized, but some major ones that are unquestionably very common and influential in familiar thinking on all subjects have been emphasized in logic textbooks and manuals since the time of Aristotle - see Hamblin (1970) for a survey of these and their historical background. Fallacies were first systematically studied by Aristotle in his book *De Sophisticis Elenchis* (*On Sophistical Refutations*) and many of the textbook treatments have basically followed Aristotle's, except that more fallacies have been added along the way. Below, some of the traditionally most important fallacies are briefly characterized.

The '*ad hominem*' fallacy' is the use of personal attack to try to undermine or refute a person's argument. Broadly, this takes two forms. The 'abusive *ad hominem*' consists of a direct attack on the arguer's character, and in particular, honesty. The 'circumstantial *ad hominem*' alleges some form of conflict between the arguer's personal circumstances and the argument. For example, it might be the charge that the arguer 'does not practise what he preaches', or that the arguer is biased because of having something to gain financially by advocating the argument in question.

The '*ad verecundiam*' fallacy' is the use of appeal to expert opinion to support one's argument in such a way as to take advantage of one's opponent's respect for authority, allowing no room to question the pronouncement of an expert. Appeal to expert opinion is a common, and in principle legitimate, type of argumentation, but one that is easily abused by being pressed too heavily in an argumentative exchange. Experts can be (spectacularly) wrong in some cases, and their views can be misquoted, or otherwise distorted, especially when they use technical language.

The '*ad ignorantiam*' fallacy' is the exploitation of ignorance (lack of knowledge) by one party in a dialogue to try to shift the burden of proof unjustifiably onto the other party. This type of argument takes two forms: (1) It is not known that proposition A is true; therefore A is false; (2) it is not known that A is false; therefore A is true. An example would be: 'This supposed theorem has never in fact been proved to be true, therefore it must be false'. In such a case, it could be that the conjecture has just been very difficult to prove, and no one has succeeded so far, but that somebody will prove it in the future. The fallaciousness of such an argument is indicated by the fact that there is such a thing as an incompleteness proof in logic; a proof that some propositions cannot be proved (see Gödel's Theorems).

The '*ad baculum*' fallacy' is the use of threat or, more broadly, scaremongering, to try to exploit the insecurity of one's opponent to get them to accept some proposition, instead of justifying the proposition by good evidence. Instances of the *ad baculum* fallacy often take the form of covert threats, for example, 'I wouldn't say that if I were you - the last person who said it wound up on the bottom of the river wearing cement boots!'

The '*ad misericordiam*' fallacy' is the blowing up out of proportion of an appeal to pity to cover up for lack of evidence to support a claim; or to brush aside other considerations that should be given more importance in arriving at a conclusion. There are many instances of quite reasonable arguments based on appeals to compassion or sympathy, but it is easy to blow such emotional appeals out of proportion to the slender weight of evidence that should be attached to them (within a larger evidential picture).

The '*ad populum*' fallacy' is committed when an arguer suggests that someone's argument is wrong or unacceptable in so far as it deviates from popular opinion, or the conventionally accepted view. When pursued with rhetorical enthusiasm, this type of argumentation is called 'mob appeal'.

'Begging the question' is a kind of fallacious circular reasoning in which the conclusion is itself required as a premise to support the argument being advanced to justify the conclusion. Many cases of this type of fallacy are

disguised by being quite complex arguments - see Walton (1991) - but a simple case would be the following argument: someone asked to prove that God is benevolent replies, 'God has all the virtues, therefore God is benevolent.' Since benevolence is presumably a virtue, the premise requires the prior assumption that God is benevolent.

'Equivocation' is the fallacy whereby an argument contains an ambiguous word or phrase that means one thing in one premise, but something different in another premise or in the conclusion. As a consequence of this shift of meaning, one may be fooled into thinking one has been presented with a single argument, which looks pretty good.

The 'straw man fallacy' is committed when one arguer misrepresents another's position to make it appear less plausible than it really is, in order more easily to criticize or refute it. For example, an environmentalist position may be exaggerated by saying, 'They are trying to make this country a paradise on earth, by barring all industrial development. They will cause all of us to become unemployed.'

The 'fallacy of affirming the consequent' appears to be valid because of its superficial similarity to the valid form of argument *modus ponens*. An example of this valid form would be: 'If it is raining, the streets are wet. It is raining. Therefore, the streets are wet.' The comparable argument in the form of affirming the consequent may have the superficial appearance of also being valid: 'If it is raining, the streets are wet. The streets are wet. Therefore, it is raining.' (But maybe the streets are wet because a water main has burst.) It is perhaps because of our tendency to reverse the first premise in some instances that we can be deceived by this kind of fallacious reasoning.

See also: [Logic, ancient §1](#); [Logical and mathematical terms, glossary of](#); [Formal and informal logic](#)

DOUGLAS WALTON

References and further reading

Aristotle (c.330 BC) *De Sophisticis Elenchis*, trans. E.S. Forster, *On Sophistical Refutations*, Loeb Classical Library Edition, Cambridge, MA: Harvard University Press, 1938.(Historical.)

Hamblin, C.L. (1970) *Fallacies*, London: Methuen; repr. Newport News, VA: Vale Press, 1993.(Still extremely useful, especially on historical aspects.)

Walton, D.N. (1991) *Begging the Question*, New York: Greenwood Press.(On circular arguments.)

Walton, D.N. (1995) *A Pragmatic Theory of Fallacy*, Tuscaloosa, AL: University of Alabama Press.(Up to date and comprehensive.)

Forcing

The method of forcing was introduced by Paul J. Cohen in order to prove the independence of the axiom of choice (AC) from the basic (ZF) axioms of set theory, and of the continuum hypothesis (CH) from the accepted axioms (ZFC = ZF + AC) of set theory (see set theory, axiom of choice, continuum hypothesis). Given a model M of ZF and a certain $P \in M$, it produces a 'generic' $G \subseteq P$ and a model N of ZF with $M \subseteq N$ and $G \in N$. By suitably choosing P , N can be 'forced' to be or not be a model of various hypotheses, which are thus shown to be consistent with or independent of the axioms. This method of proving undecidability has been very widely applied. The method has also motivated the proposal of new so-called forcing axioms to decide what is otherwise undecidable, the most important being that called Martin's axiom (MA).

1 The method: three basic lemmas

Let M be a countable transitive model of ZFC (see [Set theory §7](#), terminology and notation from which will be used throughout). Let $P \in M$. Then $p, q \in P$ are called *compatible* if there is an $r \in P$ with $p \subseteq r$ and $q \subseteq r$. $D \subseteq P$ is *dense* if it is nonempty and for every $p \in P$ there exists a $q \in D$ with $p \subseteq q$. If F is a set of dense sets, G is *generic* for F if: (1) $q \in G$ whenever $q \subseteq p$ and $p \in P$, (2) any $p, q \in G$ are compatible, and (3) for any $D \in F$, $G \cap D \neq \emptyset$. G is *generic* for M if G is generic for $F = \{D \subseteq P \mid D \text{ dense and } D \in M\}$ (which appears uncountable to M but is really countable). If $F = \{D_0, D_1, D_2, \dots\}$ is countable, choosing $p_0 \subseteq p_1 \subseteq p_2 \subseteq \dots$ with $p_k \in D_k$, as is possible since each D_k is dense, $G = \{q \in P \mid q \subseteq p_k \text{ for some } k \in \omega\}$ is generic. A generic G generally cannot be an element of M .

For example, let P be the set of all functions p with domain a finite subset of ω and range a subset of 2. (Recall that an ordinal is identical with the set lesser ordinals, so $2 = \{0, 1\}$, $\omega = \{0, 1, 2, \dots\}$.) Then p, q are compatible unless they 'disagree' in the sense that there is an $n \in \text{dom}(p) \cap \text{dom}(q)$ with $p(n) \neq q(n)$. For each $n \in \omega$, $D_n = \{p \in P \mid n \in \text{dom}(p)\}$ is dense, since any function p not already defined on n can be extended to a function q that is. Likewise, for any function $f \in M$ from ω to 2, $D_f = \{p \in P \mid p(n) \neq f(n)\}$ is dense, since any function p with a finite domain that does not already disagree with f can be extended to a function that q does. In the example, if all the $D_n, D_f \in F$ for all n and f as above, then $g = \cup G$ will be a function g from ω to 2 disagreeing with each $f \in M$. Hence $g \notin M$.

Cohen's method associates to any given countable transitive model M of ZFC, and any given $P \in M$, and any given $G \subseteq P$ generic for M another countable transitive model $N = M[G]$ of ZFC, with $M \subseteq M[G]$ and $G \in M[G]$, in such a way that three basic lemmas stated below hold. For various choices of P the three basic lemmas can be used to show that $M[G]$ will be a model of various hypotheses Δ . Before stating the three basic lemmas it should be remarked that the best way to gain an intuitive understanding of their rather technical content is to see how they are used in applications (as in §§2,3).

Minimality lemma The $b \subseteq M$ with $b \in M[G]$ will be precisely those $b \subseteq M$ of the form $\delta_G(c) = \{x \mid (p, x) \in c \text{ for some } p \in G\}$ for some $c \in M$. Note that such a $\delta(c)$ would certainly have to belong to any countable transitive model N with $M \subseteq N$ and $G \subseteq N$. It is sometimes said that while M cannot 'know' about which p will get into G or about which $b \subseteq M$ will get into $M[G] - M$ (since G and such b simply do not exist in M), M does contain 'codes' c for such b , which can immediately be 'decoded' by δ_G once G is given. Especially important are the *canonical* code $x^\# = P \otimes x$ for any $x \in M$, and the *canonical* code $\Gamma = \{(p, p) \mid p \in P\}$ for the generic set. It is fairly easily checked that for any generic G , $\delta_G(x^\#) = x$ and $\delta_G(\Gamma) = G$.

Truth lemma If $\Phi(\delta_G(a), \delta_G(b), \dots)$ holds inside $M[G]$, then there is some $p \in G$ such that $\Phi(\delta_H(a), \delta_H(b), \dots)$ holds inside $M[H]$ for any H generic for M with $p \in H$. Such a p is said to *force* $\Phi(a, b, \dots)$ over M .

Definability lemma To every formula Φ there may be associated a formula $\Phi^\dagger$ such that given any $p \in P$ and any $a, b, \dots \in M$, then $\Phi^\dagger(p, a, b, \dots)$ holds in M if and only if p forces $\Phi(a, b, \dots)$ over M . Note that it follows, by the axiom of separation, that for any set A that is an element of M , the set $\{(p, a, b, \dots) \mid u, v, \dots \in A \text{ and } p \text{ forces } \Phi(a, b, \dots)\}$ is an element of M . It is sometimes said that though M

cannot ‘know’ which p will get into G or whether a given Φ will come to hold in $M[G]$, at least Φ will come to hold in $M[G]$ only if it is forced to hold by some p that gets into G , and M ‘knows’ which p will, if they get into G , force Φ to hold in $M[G]$.

2 Application to the continuum hypothesis

To treat CH, let now ξ, η be the ordinals inside M that appear to M to be ω_1, ω_1 (really, they are, like M itself, countable), and vary the example of §1 by letting P be the set of all functions p with domain a finite subset of $\eta \otimes \omega$ and range a subset of 2. If G is generic, $g = \cup G$ will be a function from $\eta \otimes \omega$ to 2, and there will be functions g_ρ from ω to 2 for $\rho < \eta$, given by $g_\rho(n) = g(\rho, n)$. Since for distinct $\rho, \sigma < \eta$, $D_{\rho, \sigma} = \{p \in P \mid (\rho, n), (\sigma, n) \in \text{dom}(p) \text{ and } p(\rho n) \neq p(\sigma n)\}$ is dense, these functions will be distinct not only from any function in M , but from each other. So $\sim\text{CH}$ will hold inside $M[G]$ provided ξ, η still appear to $M[G]$ to be ω_1, ω_2 .

The proviso means there is no function e or f in $M[G]$ with $\text{dom}(e) = \omega$, $\xi \subseteq \text{ran}(e)$ or $\text{dom}(f) = \xi$, $\eta \subseteq \text{ran}(f)$. Cohen showed that for this it suffices that P should appear to M to have a certain property, the *countable chain condition* (CCC). Here an *antichain* is an $A \subseteq P$ such that p, q are incompatible for any two $p, q \in A$, and CCC is the property that any antichain is countable. (It ought logically to be called the ‘no countable antichain condition’, but the illogical terminology is well established.) To sketch why this suffices to rule out the existence of an e as above (the case of an f as above being similar), suppose $e \in M[G]$ is a function with $\text{dom}(e) \subseteq \omega$.

By the minimality lemma, $e = \delta_G(c) \in M$ for some $c \in M$. By the truth lemma there is a $p_0 \in G$ forcing ‘ c is a function with $\text{dom}(c) = \omega^\#$ and $\xi^\# \subseteq \text{ran}(c)$ ’, and by restricting attention to those $p \in P$ such that $p_0 \subseteq p$, we may suppose that every $p \in P$ forces this. By the definability lemma $E = \{(p, n, \rho) \in P \otimes \omega \otimes \xi \mid p \text{ forces ‘the value of function } c \text{ for argument } n^\# \text{ is } \rho^\#’}\} \in M$.

Now consider how things appear inside M , where P appears CCC. For each $n \in \omega$ let $C_n = \{\rho < \xi \mid (p, n, \rho) \in E \text{ for some } p \in P\}$, and let $C = \cup\{C_0, C_1, C_2, \dots\}$. Intuitively, C_n is the set of ‘candidates’ to become the output of the function for input n , while C is the set of ‘candidates’ to belong to the range of the function. For each $n \in \omega$ and $\rho \in C_n$ choose a $p(n, \rho)$ with $(p(n, \rho), n, \rho) \in E$, and let $A_n = \{p(n, \rho) \mid \rho \in C_n\}$. Note that if $(p, n, \rho), (q, n, \sigma) \in E$ and $\rho \neq \sigma$, then p, q are incompatible (since they force things which cannot both hold in $M[G]$ for any G). Hence any two elements of any A_n are incompatible. By CCC, each A_n , hence each C_n and hence C are all countable. Since $\xi = \omega_1$ (or so it appears inside M), C is bounded by some $\sigma < \xi$. Back outside M this means that $e = \delta_G(c)$ will be a function with $\text{ran}(e) \subseteq \sigma < \xi$, contrary to assumption that $\text{ran}(e) = \xi$.

To show that the set of functions p with domain a finite subset of $\omega_2 \otimes \omega$ and range a subset of 2 (which set is what P appears inside M to be) is CCC, the *delta system theorem* from combinatorial set theory is used (see [Set theory §4](#)): given an uncountable set A of incompatible elements of P , there must be a finite b and an uncountable $B \subseteq A$ such that $\text{dom}(p) \cap \text{dom}(q) = b$ for all $p, q \in B$. Thus the restrictions of the $p \in B$ to b give uncountably many distinct functions from b to 2. But if k be the number of elements in b , the number of such functions is just 2^k , which is finite, a contradiction.

3 Other applications

Among many other uses of the method of forcing, the following may be cited: If CH fails, the number of cardinals between $\aleph_0$ and c can be almost anything from one (hypothesis CH’) to c (hypothesis CH[#]). Moreover, as regards the function $\exp \alpha = 2^\alpha$, all kinds of combinations are possible: One can have $\exp \aleph_0 = \exp \aleph_1 = \aleph_2$, or one can have $\aleph_0 = \aleph_2$ and $\exp \aleph_1 = \aleph_3$, and so on.

Cohen himself proved the independence of AC from ZF, combining forcing with an earlier construction used to prove independence of AC from a variant of ZF allowing ‘individuals’. Indeed, Cohen’s proof established the independence of the weaker axiom of dependent choice DC. The independence of AC from ZF+DC can similarly be established. Indeed, a whole range of implications of AC have been studied and questions as to which imply which and which are independent of which have been settled.

Many undecidability results have also been obtained by forcing in combinatorial set (see [Set theory §4](#)). For instance, any combination of CH or $\sim$ CH with SH or $\sim$ SH is consistent, where SH is Suslin's hypothesis.

Many undecidability results have also been obtained by forcing in the related areas of descriptive set theory and large cardinal theory (see [Set theory §§5,7](#)). In particular, adding the hypotheses IC, MC, ... of the existence of inaccessible, measurable, ... cardinals to ZFC does not decide CH (or SH or any other such combinatorial hypothesis).

As for the hypothesis AM (resp. PM) that all (resp. all projective) sets of reals are Lebesgue measurable, in a noteworthy early extension of Cohen's work, Robert Solovay established consistency relative to ZF+IC for DC+AM and AC+PM. (Also projective uniformization PU fails in one of the models involved.) For the related hypothesis RM (resp. RSM) that there is a measure with all the properties of Lebesgue measure except invariance (resp. with all the properties of Lebesgue measure except invariance, plus superadditivity), Solovay proved the consistency of RSM (which is known to imply CH[#]) relative to ZFC+MC, and of MC relative to ZFC+RM (which is known to imply $\sim$ CH and $\sim$ CH').

In many applications, it is convenient to work with a generalization of the notions of §1: one starts with an arbitrary set P and an arbitrary partial order $\leq$ on it, and defines p, q to be compatible if there is an $r \in P$ with $p \leq r$ and $q \leq r$, $D \subseteq P$ to be dense if it is nonempty and for every $p \in P$ there exists a $q \in D$ with $p \leq q$, and so on. In §1 only the special case where $\leq$ is the subset relation $\subseteq$ was considered, but the general case easily reduces to this special case, since setting $[p] = \{r \in P \mid r \leq p\}$ one has $[p] \subseteq [q]$ if and only if $p \leq q$.

4 Martin's axiom

D.A. Martin and Solovay, by a method involving reduction of several successive applications of forcing by $P \in M$, $Q \in M[G]$, $R \in M[G][H]$... to a single application, showed the consistency with $\sim$ CH (indeed with CH') of a consequence of CH called *Martin's axiom* (MA). MA is the most basic example of a *forcing axiom*, asserting that anything of a certain kind that could be forced to hold already does hold. MA asserts that for any P that is CCC and any set F of fewer than continuum many (c) dense subsets of P , there exists a G generic for F . (Assuming CH, c amounts to $\leq \aleph_0$, and the existence of G in this case was essentially proved in §1.)

To give at least one example, consider the usual order on the reals $\mathbb{R}$ (as in [Set theory](#), end of §4, beginning of §5): $S \subseteq \mathbb{R}$ is *open dense* if for every rational a, c with $a < c$ there are rational d, e with $a \leq d < e < c$ such that $b \in S$ for every real b with $d \leq b \leq e$. It is easily proved that the intersection of finitely many open dense sets is open dense and hence not empty. It is an important theorem of modern analysis and topology that the intersection of countably many open dense sets is not empty. It is an interesting consequence of MA that an intersection of fewer than continuum many open dense sets is not empty.

To derive this consequence, let X be a set of fewer than continuum many open dense sets, to prove $\bigcap X \neq \emptyset$. Consider the set P of triples (a, c, Y) where $Y \subseteq X$ is finite, where a, c are rational with $a < c$, and where $b \in \bigcap Y$ for every real b with $a \leq b \leq c$. Consider the partial order on P given by $(a, c, Y) \leq (d, e, Z)$ if and only if $Y \subseteq Z$ and $a \leq d < e \leq c$. Then (a, c, Y) and (a', c', Y') are compatible if and only if $\max(a, a') < \min(c, c')$. For then letting $Z = Y \cup Y'$, using the fact that $\bigcap Z$ is open dense, there are rational d, e with $\max(a, a') \leq d < e \leq \min(c, c')$ such $b \in \bigcap Z$ for every real b with $d \leq b \leq e$, and so $(a, c, Y), (a', c', Y') \leq (d, e, Z)$.

In particular, (a, c, Y) and (a', c', Y') are always compatible if $d = d'$ and $e = e'$, and since there are only countably many possibilities for d, e , any set of pairwise incompatible elements of P must be countable, and the partial order is CCC. Also, for any $(a, c, Y) \in P$ and any $S \in X$, letting $Z = Y \cup \{S\}$ and using the fact that $\bigcap Z$ is open dense, there are rational d, e with $a \leq d < e \leq c$ such that $b \in \bigcap Z$ for all real b with $d \leq b \leq e$, and so $(a, c, Y) \leq (d, e, Z)$ and $S \in Z$. Thus for each $S \in X$ the set $Q_S = \{(d, e, Z) \mid S \in Z\}$ is dense in P . A similar argument shows that for each $n = 1, 2, 3, \dots$, the set $R_n = \{(d, e, Z) \mid d - e < 1/n\}$ is dense in P . Applying MA, there is a $G \subseteq P$ generic for $\{Q_S \mid S \in X\} \cup \{R_n \mid n = 1, 2, 3, \dots\}$.

Let α be the least upper bound of $\{a \mid (a, c, Y) \in G \text{ for some } c, Y\}$ and let γ be the greatest lower bound of $\{c \mid (a, c, Y) \in G \text{ for some } a, Y\}$. Since any elements $(a, c, Y), (a', c', Y') \in G$ are compatible, and so have $\max(a, a') < \min(c, c')$, one has $\alpha \leq \gamma$. For any n there is an $(a, c, Y) \in G \cap R_n$, whence

$(\alpha - \gamma) \leq (a - c) < 1/n$, and since this holds for all $n = 1, 2, 3, \dots$, one has $\alpha = \gamma$. For any $S \in X$ there is an $(a, c, Y) \in G \cap Q_s$, and since $a \leq \alpha \leq \gamma$ one has $\alpha \in S$, and since this holds for all $S \in X$, one has $\alpha \in \cap X$, completing the proof.

Of the most interesting consequences of CH, about half follow from MA alone, about half are refutable assuming $\text{MA} + \sim\text{CH}$, which also has other interesting consequences ($\sum_2^1$ -measurability, SH). By contrast, $\sim\text{CH}$ by itself has few interesting consequences. $\text{MA} + \sim\text{CH}$ thus provides a rather attractive alternative to CH, though there is no generalization of $\text{MA} + \sim\text{CH}$ to higher cardinals comparable to the generalized continuum hypothesis (GCH) generalizing CH. Saharon Shelah has introduced forcing axioms considerably strengthening $\text{MA} + \sim\text{CH}$ and implying CH'.

5 On the proofs of the three basic lemmas

Only the briefest sketch of the highly technical proofs of the three main lemmas may be given here. It cannot be emphasized strongly enough that the best way to gain an intuitive understanding of the three main lemmas is to see how they are used in applications (as in §§2,3), which hardly ever refer back to details of their proofs.

First, while the notion of coding given in §1 is that generally needed in applications, for the proofs a different notion of coding is needed, one providing codes for *all* elements of $M[G]$ and not just for those that are subsets of M . Correspondingly, a different notion of decoding is needed: $\delta_G(c) = \{\delta_G(b) \mid (p, b) \in c \text{ for some } p \in G\}$. This definition appears circular but is really an instance of a legitimate method of definition of operations on sets by $\in$ -recursion, a generalization of the method of definition of operations on ordinals by $<$ -recursion (or transfinite recursion, as in [Set theory §4](#)): As an operation on ordinals may be defined by defining its value for ρ in terms of its values for $\sigma < \rho$, so an operation on sets may be defined by defining its value for c in terms of its values for $b \in c$. With this alternate definition of decoding, $M[G]$ is simply defined to be $\{\delta_G(c) \mid c \in M\}$, and to establish the minimality lemma what remains to be proved is that $M[G]$ so defined is a model of ZFC. To prove this, one makes heavy use of various necessary and sufficient conditions for being a model of this or that axiom of set theory (see [Set theory §7](#)). For instance, to handle the axiom of pairing, one must show that $M[G]$ is closed under the operation of pairing: One must show that if $a, b \in M$, then there is a $c \in M$ such that $\delta_G(c) = \{\delta_G(a), \delta_G(b)\}$. In this case it is easily checked that the required c is provided by $P \otimes \{a, b\}$.

Second, while the definition of forcing given in §1 is that generally needed in applications, for the proofs a different definition is needed initially, though the two definitions will be proved equivalent *eventually*. This alternate definition of p forces Φ is by induction on the complexity of the formula Φ : (i) p forces 'not Φ ' if and only if no q with $p \subseteq q$ forces Φ ; (ii) p forces ' Φ and Ψ ' if and only if p forces Φ and p forces Ψ ; (iii) p forces 'for all x , $\Phi(x)$ ' if and only if for all c , p forces $\Phi(c)$; and so on. What is most difficult is to define p forces Φ for the *simplest* formulas ' $c = d$ ' and ' $c \in d$ ': The definition for these simplest formulas requires a highly technical $\in$ -recursion. With the alternate definition of forcing, the definability lemma is almost immediate.

Towards establishing the truth lemma, and with it the equivalence of the alternate definition of forcing to the definition given in §1, one notes that it follows from clause (i) of the alternate definition of forcing that the set $D = \{p \mid p \text{ forces } \Phi \text{ or } p \text{ forces } \sim\Phi\}$ is dense, and it follows from the definability lemma that $D \in M$. Hence for any generic G there will be a $p \in G \cap D$. What remains to be proved is that if $p \in G$ and p forces $\Phi(a, b, \dots)$, then $\Phi(\delta_G(a), \delta_G(b), \dots)$ holds in $M[G]$. This is proved by induction on the complexity of the formula Φ .

See also: [Logical and mathematical terms, glossary of](#)

JOHN P. BURGESS

References and further reading

- Burgess, J.** (1977) 'Forcing', in K.J. Barwise (ed.) *Handbook of Mathematical Logic*, Amsterdam: North Holland, 403-52. (A survey in a standard reference, derived from lectures of Solovay and emphasizing consistency proofs for combinatorial principles.)
- Cohen, P.J.** (1966) *Set Theory and the Continuum Hypothesis*, New York: Benjamin. (The original source for forcing, now mainly of historical interest, having been superseded in technical aspects by later, streamlined treatments.)
- Jech, T.** (1978) *Set Theory*, New York: Academic Press. (A comprehensive treatise, covering the applications

mentioned and many more besides, with attributions to their original authors and references to the original technical literature.)

Jech, T. (1986) *Multiple Forcing*, Cambridge: Cambridge University Press. (A semi-popular account of the forbiddingly technical work of Shelah.)

Martin, D.A. and Solovay, R. (1970) 'Internal Cohen Extensions', *Annals of Mathematical Logic* 2: 143-78. (The original, and a highly readable, source for Martin's axiom.)

Formal and informal logic

Formal logic abstracts the form of an argument from an instance of it that may be encountered, and then evaluates the form as being valid or invalid. The form is the important thing, rather than the concrete instance of the form. Informal logic, on the other hand, evaluates how an argument is used in a given context of conversation. This more practical, real-world orientation requires more judgment in interpreting what a given text of discourse should be taken to argue.

Formal logic abstracts from the context of use of an argument by seeing an argument as an instance of a form of argument, composed of constants and variables. The constants are defined in a precise way that makes any reference to context unnecessary (see **Logical constants**). The variables are blanks that are filled in by the component parts of the argument under consideration. Take the argument ‘Jane and Dick will not both be elected; Jane will be elected; therefore Dick will not be elected’. In the modern formal logic of propositions, ‘and’ and ‘not’ are defined as constants, and ‘Jane will be elected’ and ‘Dick will be elected’ are represented using propositional variables, *A* and *B*. The constants ‘and’ and ‘not’ are defined as follows:

‘*A* and *B*’ is true only if both *A* and *B* are true; otherwise it is false.

‘not *A*’ is true if *A* is false; and false if *A* is true.

Symbolized thus, the argument above about Jane and Dick has the following form:

not(*A* and *B*)

A

Therefore not *B*.

This form of argument is ‘valid’, meaning that every single instance of it is such that if both premises are true, the conclusion must also be true. Thus formal logic guarantees that a valid argument will never take you from true premises to a false conclusion. This is simply a matter of the form of the argument and the way the constants have been defined. Hence formal logic is an abstract, mathematical discipline - a study of forms.

Formal logic of this mathematical sort, using constants and variables, is a relatively modern development, but the roots of it can be traced back to Aristotle’s theory of the syllogism (see **Aristotle §5; Logic, ancient §§1-3**). A ‘syllogism’ is an argument made up of three propositions put together in a particular form, of which the following is an example: ‘All bears climb trees; all grizzlies are bears; therefore, all grizzlies climb trees’. This kind of argument was ruled valid in Aristotle’s theory: *if* the premises are true, then the conclusion must be true too. This does not imply that the premises *are* true. Validity, you could say, is a conditional concept in Aristotle’s theory, and the same can be said in modern logic.

Modern formal logic has been centrally concerned with two kinds of arguments. Propositional logic has to do with arguments that are valid or not by virtue of certain constants, such as ‘and’, ‘not’ and ‘if..., then...’, which function as connectives between propositions. Quantifier logic is a sophisticated and powerful system of formal logic that not only encompasses the limited varieties of arguments treated in Aristotle’s theory of the syllogism, but also many other more complex arguments depending on ‘all’ and ‘some’ (see **Quantifiers; Predicate calculus**).

Informal logic was also founded as a systematic subject by Aristotle, principally in his *Topics* and *Sophistical Refutations* (see **Topica et Sophistici Elenchi**), but never received the same consistent and more serious attention that formal logic did in the subsequent history of logic. Consequently, it was never developed as a scientific subject in the way that formal logic was. However, it survived in a practical form, in logic textbooks and manuals used for instruction in universities throughout the Middle Ages and into the modern period. Only recently, with the advent of argumentation theory, has there come to be a relatively serious and sustained scholarly attempt to develop informal logic as a subject of research. Yet even at the end of the twentieth century, informal logic is a more practical subject which addresses contextual factors of how arguments are actually used in everyday conversation in a given discourse. Unlike formal logic, informal logic cannot abstract from the context of use of an argument, and must evaluate it as subject to an interpretation of the speaker’s discourse as a part of a conversational exchange.

Both formal and informal logic are concerned primarily with the task of evaluating arguments as correct or incorrect, according to standards. Formal logic, as noted above, works by identifying forms and proving them valid or invalid by means of some method of calculation independent of the context in which the argument was used. Applying the formal test to a real argument does involve some informal determining of what kind of argument was meant by the speaker, what its most precise form is, and so forth. But in informal logic such questions of interpretation are a much more central part of the task of evaluation.

Informal logic has to do with the use of an argument in a conversational setting where participants are supposed to contribute to the conversation by making moves appropriate to the current stage of the conversation (see [Grice 1975](#); [Pragmatics](#); [Implicature](#)). Different types of conversation can be recognized. For example, van Eemeren and Grootendorst (1984) cite one very important type where the purpose is to resolve a conflict of opinions by verbal means, using arguments. Generally an argument is correct if it contributes to the goal of the conversation, and incorrect if it fails to do so. Some particularly deceptive or misleading types of arguments can even actively interfere with the goal of a conversation, and the most common and dangerous of these are called fallacies (see [Fallacies](#)).

See also: [Logical and mathematical terms, glossary of](#)

DOUGLAS WALTON

References and further reading

- Aristotle** (c.330 BC) *Topica et Sophistici Elenchi (Topics and Sophistical Refutations)*, trans. W.A. Pickard-Cambridge, ed. W.D. Ross, New York: Oxford University Press, 1958.(Historical.)
- Eemeren, F.H. van and Grootendorst, R.** (1984) *Speech Acts in Argumentative Discussions*, Dordrecht: Foris. (Account of argumentation theory.)
- Grice, H.P.** (1975) 'Logic and Conversation', in D. Davidson and G. Harman (eds) *The Logic of Grammar*, Encino, CA: Dickenson.(Account of implicature.)
- Jeffrey, R.C.** (1967) *Formal Logic: Its Scope and Limits*, New York: McGraw-Hill, 2nd edn, 1981.(General introduction to formal logic.)
- Walton, D.** (1989) *Informal Logic*, Cambridge: Cambridge University Press.(General introduction to informal logic.)

Formal languages and systems

Formal languages and systems are concerned with symbolic structures considered under the aspect of generation by formal (syntactic) rules, that is, irrespective of their or their components' meaning(s). In the most general sense, a formal language is a set of expressions. The most important way of describing this set is by means of grammars. Formal systems are formal languages equipped with a consequence operation yielding a deductive system. If one further specifies the means by which expressions are built up (connectives, quantifiers) and the rules from which inferences are generated, one obtains logical calculi of various sorts, especially Frege-Hilbert-style and Gentzen-style systems.

1 Expressions and grammars

A formal language is a set $\mathcal{L}$ of expressions. Expressions are built up from a finite set Σ of atomic symbols or atoms (the 'alphabet' of the language) by means of certain constructors. Normally one confines oneself to linear association (concatenation) ' $\circ$ ' as the only constructor, yielding 'strings' of atoms, also called 'words' (over Σ), starting with the empty word ϵ . Mathematically, the structure obtained is the semi-group freely generated by the alphabet Σ . When $a_1 \circ a_2 \circ \dots \circ a_n$ with atoms $a_1, a_2, \dots, a_n$ is written in the usual way as $a_1 a_2 \dots a_n$, it is assumed that the decomposition of this expression into atoms is unique.

It should be noted that languages based on more than one constructor may well be considered. An example would be Frege's two-dimensional *Begriffsschrift*, which may be viewed as a formal language based on constructors for horizontal and vertical alignment (see Frege 1879). However, since any constructor can be written linearly as a function symbol applied to arguments, the notion of formal languages as based on linear concatenation is sufficiently general. The atomic symbols which form the basis of a formal language are themselves abstract entities, in contrast to the tokens which constitute their individual realizations.

A formal language $\mathcal{L}$ can be investigated from various points of view. One possibility is to specify a device $\mathcal{A}$ that recognizes exactly those words over Σ which belong to $\mathcal{L}$, that is, which enters a certain state if and only if a word of $\mathcal{L}$ is given as an input. When this occurs we say that $\mathcal{A}$ *accepts* $\mathcal{L}$. Such a device is called an 'automaton' (in the mathematical, not the physical sense). The relationship between languages and automata accepting them is studied in automata theory.

A formal grammar $\mathcal{G}$ for $\mathcal{L}$ describes a way to generate the words of $\mathcal{L}$ by means of formal rules, beginning with a special start symbol S . A rule (also called a 'production') has the form $u \rightarrow v$, which expresses that in any word of the form $w_1 u w_2$ the part u may be replaced by v , yielding $w_1 v w_2$. In addition to the atoms of the alphabet Σ under consideration (which are called 'terminals' in formal language theory), rules may contain variables (called 'non-terminals'), among which S is distinguished. For example, the grammar

$$S \rightarrow ASA$$

$$S \rightarrow a$$

$$A \rightarrow b$$

generates all words $b \dots bab \dots b$ over the alphabet $\{a, b\}$ with equally many instances of b on each side of a .

Formal languages may be classified by the type of grammar generating them. For example, a language $\mathcal{L}$ is called 'context-free' if it can be generated by a context-free grammar, which is one whose productions always have just a single variable on the left side (as in the example just given). Context-free grammars play a distinguished role in the construction of artificial languages (such as programming languages). However, for the study of (fragments of) natural languages, more complicated ('context-sensitive') types of grammar must be considered. Formal grammars have the same expressive power as computable functions: the computation by a Turing machine can be described by means of a formal grammar; and the generation of words in a formal grammar can be simulated by a Turing machine.

2 Deductive systems

A formal system is based on a formal language $\mathcal{L}$, endowing it with a 'consequence operation' $\mathcal{C}$. This operation

can be specified at different levels of abstraction. In the most general sense, $\mathcal{C}$ is just an arbitrary function transforming subsets of $\mathcal{L}$ into subsets of $\mathcal{L}$. $\mathcal{C}$ is said to be an ‘inference operation’ if the set of consequences of a set X comprises at least X .

$$X \subseteq \mathcal{C}(X) \text{ (inclusion)}$$

The pair $\langle \mathcal{L}, \mathcal{C} \rangle$ is then called an ‘inference system’. It is called a ‘closure system’ if furthermore

$$\mathcal{C}(\mathcal{C}(X)) \subseteq \mathcal{C}(X) \text{ (idempotence)}$$

and

$$X \subseteq Y \Rightarrow \mathcal{C}(X) \subseteq \mathcal{C}(Y) \text{ (monotonicity)}$$

are fulfilled. It is called a ‘deductive system’ if the consequences of a set X can be obtained from the finite subsets of X , that is, if in addition to the three conditions mentioned,

$$\mathcal{C}(X) \subseteq \bigcup \{ \mathcal{C}(Y) : Y \subseteq X, Y \text{ finite} \} \text{ (compactness)}$$

holds. Equivalently, formal systems can be described by a consequence relation $X \vdash A$ between subsets of $\mathcal{L}$ and expressions of $\mathcal{L}$. The four conditions mentioned then become

$$\begin{aligned} X \cup \{A\} &\vdash A \\ X \vdash A &\Rightarrow X \cup Y \vdash A \\ (X \vdash A \text{ for all } A \in Y \text{ and } Y \cup Z \vdash B) &\Rightarrow X \cup Z \vdash B \\ X \vdash A &\Rightarrow Y \vdash A \text{ for some finite } Y \subseteq X. \end{aligned}$$

If we confine ourselves to consequences of finite sets X, Y, Z , which in the case of deductive systems is appropriate, these conditions are equivalent to

$$\text{to } \{A\} \vdash A \text{ (identity)} \quad X \vdash A \Rightarrow X \cup \{B\} \vdash A \text{ (thinning)} \quad (X \vdash A \text{ and } \{A\} \cup Z \vdash B) \Rightarrow X \cup Z \vdash B \text{ (cut)}$$

which are the basic structural (that is, logic-free) principles of Gentzen-style sequent systems (see [Gentzen 1935](#)).

Deductive systems can be given by means of a set Δ of inference rules R , where an inference rule is an $(n+1)$ -ary relation $R \subseteq \mathcal{L}^n \times \mathcal{L}$. Any $(A_1, \dots, A_n, B) \in R$ (also written as $A_1, \dots, A_n \Rightarrow B$) is called an instance of R , and $A_1, \dots, A_n$ are said to be the premises and B the conclusion of that instance. If $n = 0$ then the rule $\Rightarrow A$ is called an axiom. If we define $\mathcal{C}_\Delta(X)$ to be the smallest set of expressions in $\mathcal{L}$ containing X and closed under the rules in Δ (that is, if $\{A_1, \dots, A_n\} \subseteq \mathcal{C}_\Delta(X)$, then $B \in \mathcal{C}_\Delta(X)$ for any instance $A_1, \dots, A_n \Rightarrow B$ of a rule in Δ), then $\langle \mathcal{L}, \mathcal{C}_\Delta \rangle$ is a deductive system, called the rule-based system with respect to Δ .

Conversely, the consequence relation of any deductive system defines a set of rules with respect to which this system can be viewed as a rule-based system. Like grammars, rule-based systems are as powerful as Turing machines: they can be used to express any algorithm.

3 Logical calculi

In logic one considers deductive systems over a language $\mathcal{L}$ whose expressions, called formulas, are built up from certain basic expressions by means of functions, predicates and logical operators. In propositional logic, which for reasons of simplicity is considered here, formulas are built up from a subset $\mathcal{P}$ of $\mathcal{L}$ (the set of propositional letters or propositional variables) and a finite set $\mathcal{S}$ of symbols of the alphabet Σ of $\mathcal{L}$ (the set of propositional connectives). Prominent connectives are ‘not’ ($\sim$), ‘and’ ($\wedge$), ‘or’ ($\vee$) and ‘implies’ ($\supset$). The two principal ways of specifying a consequence operation $\mathcal{C}$ for such a system are (1) by presenting it as a rule-based system and (2) by formally describing its consequence relation $\vdash$. Other approaches modify or extend the notion of consequence by considering (3) multiple consequences, (4) consequence relations with restricted structural principles (‘substructural logics’) or (5) the fine structure of proofs.

(1) For a *rule-based system* a set of axioms and a set of inference rules are given. Axioms for a fragment of propositional logic based on conjunction and implication (the so-called ‘positive implication-conjunction logic’) are, for example, all formulas of the form

$$\begin{aligned}
 &A \supset (B \supset A) \\
 &(A \supset (B \supset C)) \supset ((A \supset B) \supset (A \supset C)) \\
 &(A \supset (B \supset C)) \supset ((A \wedge B) \supset C) \\
 &((A \wedge B) \supset C) \supset (A \supset (B \supset C)).
 \end{aligned}$$

Inference rules are all instances of the following schema.

$$A, A \supset B \Rightarrow B (\textit{modus ponens})$$

Rule-based systems of this kind are often called Frege-Hilbert-style systems. Typically, they include a relatively long list of axioms and just a few inference rules (in ordinary propositional logic, just *modus ponens*). The consequence relation ‘ $X \vdash_{\mathcal{C}} A$ ’ corresponding to the consequence operator $\mathcal{C}$ is called ‘derivability from assumptions’.

(2) In contradistinction to this approach, Gentzen-style sequent systems formulate the desired properties of a consequence relation ‘ $X \vdash_{\mathcal{C}} A$ ’ as formal rules of inference. In a sequent system of propositional logic based on the connectives ‘ $\wedge$ ’ and ‘ $\supset$ ’ the above rules of identity, thinning and cut are typically postulated as logic-free structural rules, in addition to the following logical rules.

to $X \cup \{A\} \vdash B \Rightarrow X \vdash A \supset B$ ($\supset$ -introduction right) $X \vdash A, X \cup \{B\} \vdash C \Rightarrow X \cup \{A \supset B\} \vdash C$ ($\supset$ -introduction left) $X \vdash A, X \vdash B \Rightarrow X \vdash A \wedge B$ ($\wedge$ -introduction right) $X \cup \{A, B\} \vdash C \Rightarrow X \cup \{A \wedge B\} \vdash C$ ($\wedge$ -introduction left)

Formally, this is just a specification of a rule-based system with respect to a more complicated language. Rather than a set of formulas $\mathcal{L}$ over $\mathcal{P}$ and $\mathcal{S}$ one considers a language $\mathcal{L}'$, whose expressions are ‘sequents’ of the form ‘ $X \vdash A$ ’, where $A \in \mathcal{L}$ with X finite. The deductive system $\mathcal{C}'$ over $\mathcal{L}'$ is then given by the rules stated. (Actually, to capture the idea of finite *sets* X in *expressions* ‘ $X \vdash A$ ’, one has to consider lists instead and to add special inference rules.) Intuitively, $\langle \mathcal{L}', \mathcal{C}' \rangle$ states features of the relation $\vdash_{\mathcal{C}}$ which are intended to describe properties of the connectives of $\mathcal{S}$. So a Gentzen-style sequent system is a rule-based system with a consequence relation $\vdash_{\mathcal{C}'}$ that can be viewed as a description of a deductive system with a consequence relation $\vdash_{\mathcal{C}}$ such that $X \vdash_{\mathcal{C}} A$ if and only if $\emptyset \vdash_{\mathcal{C}'} (X \vdash A)$.

(3) In the Gentzen tradition, more complicated types of consequence relations ‘ $\vdash$ ’ have been investigated that do not completely fit the pattern described. One such approach is to consider consequence relations with a set of formulas appearing on the right side of ‘ $\vdash$ ’ (‘multiple-conclusion logics’ or, sometimes, ‘Scott consequence relations’; see [Multiple-conclusion logic](#)). The intended reading is disjunctive: ‘ $X \vdash Y$ ’ means that under the assumptions X at least one element of Y holds (although that cannot be formally derived!). (Gentzen (1935) has proposed this idea to formulate sequent calculi for classical logic in which each connective obtains a ‘natural’ meaning via introduction rules for it on the right or left side of the turnstile without the need to consider special axioms or rules to guarantee the validity of, for example, $A \vee \sim A$ or $\sim \sim A \supset A$. Negation and disjunction rules in this system can be formulated as follows.

(Note the duality between these $\vee$ -rules and the $\wedge$ -rules formulated above, the right and left sides of the sequents being interchanged.) The *tertium non datur* (excluded middle) $\{ \vdash A \vee \sim A \}$ is then obtained from $\{A\} \vdash \{A\}$ via $\vdash \{A, \sim A\}$. It is obvious that this notion of consequence has strong elements of symmetry which are characteristic of classical logic.

(4) Another approach is to restrict the structural principles underlying deductive systems. If in $X \vdash A$ we consider X no longer as a set but as a list of formulas, we may discuss whether in all cases we want to assume that elements of X are ‘permutable’ (that is, whether $Y, A, B, Z \vdash C \Rightarrow Y, B, A, Z \vdash C$ holds), or whether two identical formulas in X may be ‘contracted’ to a single one (that is, whether $Y, A, A, Z \vdash B \Rightarrow Y, A, Z \vdash B$ holds), or whether we want to assume or reject the principle of thinning (see above).

Such considerations, which apply analogously to the multiple-conclusion case, lead to so-called ‘substructural’ logics, which have become important in computer science and linguistics (but not only there); in particular, relevance logics (logics without thinning), contraction-free logics (in which the multiplicity of formulas counts), linear logics (without thinning and contraction) and Lambek logics without thinning, contraction and permutation). If at the level of general consequence relations we consider logics without monotonicity, we enter the general field of non-monotonic logics, which is of special interest in artificial intelligence, since it helps to

frame many common forms of reasoning (see [Non-monotonic logic](#)).

(5) In proof theory (see [Proof theory](#)) one is often interested in the structure of the derivation that leads from a set of assumptions X to a conclusion A and in criteria for when two proofs of the same ‘derivability fact’ $X \vdash A$ are to be seen as identical - rather than in whether $X \vdash A$ holds or not. Significant cases are calculi of natural deduction (see [Natural deduction, tableau and sequent systems](#)). With respect to derivability, they can be described as Gentzen sequent systems in the sense above. However, crucial features of the fine structure of derivations, such as the fact that different occurrences of the same formula as an assumption may be treated differently, cannot be captured in that way.

One should be aware that in cases (3)-(5), as well as in other cases, in which extensions or modifications of the usual concept of a deductive system are considered with respect to some ‘object logic’, the system in which this object logic is described has itself the structure of a deductive system in the ordinary sense. Sequent calculi for a multiple-conclusion consequence relation ‘ $\vdash$ ’, perhaps with substructural features, are themselves rule-based systems. Even the structure of natural deduction proofs can be characterized by a rule-based system, namely a system of type assignment for λ -terms. Therefore, it is not so much the specific type of consequence relation studied which makes logical systems formal systems, but rather the fact that the variety of logical systems can be described in terms of formal systems in the very specific sense of a rule-based system.

See also: [Gentzen, G.K.E.](#); [Logical and mathematical terms, glossary of](#); [Turing machines](#)

HEINRICH HERRE
PETER SCHROEDER-HEISTER

References and further reading

- Došen, K. and Schroeder-Heister, P. (eds)** (1993) *Substructural Logics*, Oxford: Clarendon Press. (Overview of logics with restricted structural rules.)
- Frege, G.** (1879) *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*, Halle: Nebert; trans. J. van Heijenoort, ‘*Begriffsschrift, a Formula Language, Modelled Upon That of Arithmetic, for Pure Thought*’, in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic 1879-1931*, Cambridge, MA: Harvard University Press, 1967. (The first formulation of a complete axiom system for first-order logic.)
- Gentzen, G.** (1935) ‘Untersuchungen über das logische Schließen’, *Mathematische Zeitschrift* 39: 176-210, 405-565; repr. Darmstadt: Wissenschaftliche Buchgesellschaft, 1969; trans. ‘Investigations into Logical Deduction’, in *The Collected Papers of Gerhard Gentzen*, ed. M.E. Szabo, Amsterdam and London: North Holland, 1969. (Gentzen’s doctoral thesis, in which he created the notions of natural deduction and sequent calculus.)
- Harrison, M.A.** (1978) *Introduction to Formal Language Theory*, Reading, MA: Addison-Wesley. (Like Salomaa (1973), an advanced textbook of formal language theory.)
- Hopcroft, J.E. and Ullman, J.D.** (1979) *Introduction to Automata Theory, Languages and Computation*, Reading, MA: Addison-Wesley. (Elementary textbook of formal language theory.)
- Kleene, S.C.** (1952) *Introduction to Metamathematics*, Amsterdam: North Holland, 1974. (Classic textbook of mathematical logic with emphasis on logical calculi.)
- Salomaa, A.** (1973) *Formal Languages*, New York: Academic Press. (Like Harrison (1978), an advanced textbook of formal language theory.)
- Takeuti, G.** (1975) *Proof Theory*, Amsterdam: North Holland. (Textbook of Gentzen-style proof theory.)
- Tarski, A.** (1956) *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*, trans. and ed. J.H. Woodger, Oxford: Clarendon Press; repr. Indianapolis, IN: Hackett Publishing Company, 2nd edn, 1983. (Includes Tarski’s basic logical papers, particularly ‘On Some Fundamental Concepts of Metamathematics’ (1930), in which he defined the abstract notion of consequence for the first time.)
- Troelstra, A.S. and Schwichtenberg, H.** (1996) *Basic Proof Theory*, Cambridge: Cambridge University Press. (Textbook of Gentzen-style natural deduction and sequent systems.)
- Wójcicki, R.** (1988) *Theory of Logical Calculi: Basic Theory of Consequence Operations*, Dordrecht: Kluwer. (A comprehensive development of the theory of consequence relations.)

Free logics, philosophical issues in

The expression 'free logic' is a contraction of the more cumbersome 'logic free of existence assumptions with respect to both its general terms (predicates) and its singular terms'. Its most distinctive feature is the rejection of the principle of universal specification, a principle of classical predicate logic which licenses the logical truth of statements such as 'If everything rotates then (the planet) Mars rotates'. If a free logic contains the general term 'exists', this principle is replaced by a restricted version, one which licenses the logical truth only of statements such as 'If everything rotates then Mars rotates, provided that Mars exists'. If the free logic does not contain the general term 'exists', but contains the term 'is the same as', the principle is replaced by a version which licenses only statements such as 'If everything rotates then Mars rotates, provided that there is an object the same as Mars'.

Most free logicians regard the restricted version of universal specification as simply making explicit an implicit assumption, namely, that Mars exists. Indeed, free logic is the culmination of a long historical trend to rid logic of existence assumptions with respect to its terms. Just as classical predicate logic purports to be free of the hidden existence assumptions which pervaded the medieval theory of inference with respect to its general terms, so free logic rids classical predicate logic of hidden existence assumptions with respect to its singular terms.

There are various kinds of free logic, with many interesting and novel philosophical applications. These cover a wide range of issues from the philosophy of mathematics to the philosophy of religion. In addition to the issue of how to analyse singular existence statements, of the form '3 + 7 exists' and 'That than which nothing greater can be conceived exists', of special importance are issues in the theory of definite descriptions, set theory, the theory of reference, modal logic and the theory of complex general terms.

1 Kinds of free logic

Though there are various historical antecedents of free logic, it is a relatively recent development in the logic of terms - that is, in the development of predicate logic with or without identity. In fact, serious technical and philosophical study of this alternative to classical predicate logic originated only in the second half of the twentieth century with Henry Leonard's 1956 paper, '[Logic of Existence](#)'. The technical rudiments of free logic were then worked out in the subsequent decade and a half, the expression 'free logic' being first coined in 1960 (by Karel Lambert). Since then many studies discussing or employing free logic in general philosophy, philosophical logic and in the philosophy of science have appeared.

'Free logic' is short for 'logic free of existence assumptions with respect to both its general terms (predicates) and its singular terms'. It restricts the principle of universal specification (which licenses the logical truth of statements such as 'If everything rotates then (the planet) Mars rotates') to a version which licenses only statements such as 'If everything rotates then Mars rotates, provided that Mars exists' (if the logic contains the general term 'exists') or such as 'If everything rotates then Mars rotates, provided that there is an object the same as Mars' (if the free logic does not contain the general term 'exists', but contains the term 'is the same as').

When a free logic also entertains the empty universe of discourse, it is called a universally free logic. In such comprehensive free logics, statements (perhaps without singular terms) such as 'There is an object which is tall or is not tall' and 'There is an object which is self-identical' - statements which [Russell \(§§9, 11\)](#) regarded as impure logical truths because of the evidently factual implication that there exists at least one thing - are also excluded from the class of logical truths.

In any free logic, universal or otherwise, the quantifier context 'There is an object...' has existential force just as it does in classical predicate logic. So in any free logic containing the general term 'is the same as' (the identity symbol), a statement such as 'Mars exists' can be taken as shorthand for the statement 'There is an object the same as Mars'. Indeed, statements of the latter form can be false in free logic with identity, and hence, in contrast to classical predicate logic with identity, are not logically true. An example is the false statement 'There is an object the same as (the planet) Vulcan'.

Free logics may be divided into three classes depending on the treatment of atomic (or simple) statements containing at least one singular term *t* such that the statement '*t* exists' is false. For example, the statement 'Vulcan

rotates' is an atomic statement containing a singular term, 'Vulcan', such that 'Vulcan exists' is false. Those which count all such statements as false are called 'negative', those which regard some such statements as true are called 'positive', and those which assign no truth-value at all to such statements - except perhaps statements of the form 'Mars exists'(or 'Vulcan exists') - are called 'neutral'.

In virtue of these different views about the truth-value of atomic statements with singular terms that refer to no existent object there is a significant inferential difference between the different kinds of free logic. For only in negative free logic does a version of the classical principle of particularization hold. In a negative free logic, the principle of particularization licenses the logical truth of the conditional statement 'If Vulcan rotates then there is an object which rotates' because the antecedent, 'Vulcan rotates', is an atomic statement. This does not alter the falsity of the conditional 'If Vulcan is not an object then there is an object which is not an object' because its (true) antecedent, 'Vulcan is not an object', is not atomic.

The kind of free logic one adopts has epistemological significance. For instance, when Descartes' declaration 'I think, therefore I am [exist]' is construed as an argument and formulated in a positive free logic it is invalid. For when the singular term 'I' is replaced by the singular term 'Sherlock Holmes', and the general term 'think' is replaced by the general term 'is self-identical', the resulting premise 'Sherlock Holmes is self-identical' is true in virtually any positive free logic but the resulting conclusion 'Sherlock Holmes exists' is false. On the other hand, if formulated in a negative or neutral free logic, Descartes' argument is valid because the premise is either false or without truth-value when 'I' is replaced by a singular term such as 'Sherlock Holmes', no matter what general term is substituted for 'thinks'.

2 Philosophical semantics

It is a common misunderstanding that in free logics some singular terms need not refer. There do exist, however, semantic developments of free logics which entertain disjoint (and possibly empty) universes of discourse. Usually (though not always) these are philosophically interpreted as sets of existent objects and sets of nonexistent objects, after Alexius Meinong. In such developments, a singular term *always* refers, but not always to an existent object - the singular term, 'Sherlock Holmes', for example, may be taken to refer to a nonexistent (fictional) object. On the other hand, there also exist semantic treatments of free logic with a single (and possibly empty) universe of discourse. This is virtually always interpreted philosophically as the set of existent objects, after Russell. In such developments, a singular term can fail to refer to an existent object, and hence to any thing at all. Since free logics can be based on either Meinongian or Russellian ontologies, it is only correct to say that in such logics a singular term may fail to refer to an *existent* object. In free logics, successful or failed reference is explained as follows. Let 'X' be a place-holder for the name of a singular term, and let ' ' be a place-holder for the singular term named. Then a statement of the form 'X refers (to an existent object)' is true if and only if a statement of the form ' exists' is true - hence, when the language contains identity, if and only if a statement of the form 'There exists an object the same as ' is true. For example, the statement 'The satellite of the Earth refers (to an existent object)' is true if and only if the statement 'The satellite of the Earth exists' is true (or, if the language contains identity, if and only if the statement 'There is an object the same as the satellite of the Earth' is true).

The quantifiers of free logic can be interpreted either objectually or substitutionally (see **Quantifiers, substitutional and objectual**); if interpreted in the latter way, then it is necessary to say in the clause for a quantifier context of the form 'There is an object such that it ' that it is true just in case a sentence of the form 'it ' becomes true when 'it' is replaced by some name 'n' such that 'n exists' is true. This condition is needed to give the quantifier context in question the requisite existential force. In current parlance, quantifier contexts of the form 'There is an object such that it ' in free logic are 'actualist'. Treatments seeking to preserve classical predicate logic by interpreting the quantifier context in question substitutionally while nevertheless admitting singular terms that refer to no existent object are not free logics because that quantifier context has no existential force; it is, in other words, 'non-actualist' in such treatments.

3 Implications and applications

Free logics are not committed to the philosophical doctrine that existence is a 'predicate' (see **Existence**) - indeed, there are free logics without identity or an existence symbol. Neither are they committed to the philosophical doctrine that proper names are truncated definite descriptions. They are not even committed to the doctrine that the

reference of most grammatically proper names is determined by the reference of some description, a view often attributed to Frege and Russell (see **Proper names §1**). But an important result in the proof theory of positive free logic does bear on the traditional problem of defining (and, hence, explaining away) existence. Since it can be shown that, in the otherwise classical formal idiom minus identity, statements of the form ‘Mars exists’ (or ‘Vulcan exists’) are not definable in positive free logics, then, if one chooses to couch one’s logic of an existence predicate in positive free logic, that predicate cannot be eliminated save by means of the complex general term (predicate) ‘is the same as something’ or its equivalents.

Applications of free logic to important philosophical concerns are numerous. They range from the philosophy of mathematics to the philosophy of religion. Of special philosophical importance are the topics of definite descriptions, modality, the notion of reference, set theory and the theory of complex predicates.

In contrast to the Fregean tradition, which assigns an existent, sometimes arbitrarily, as referent to ‘unfulfilled’ definite descriptions such as ‘the planet Vulcan’, in contrast to the Russellian tradition which treats all definite descriptions as grammatical but not as genuine singular terms, and in contrast to the Hilbert-Bernays tradition which regards unfulfilled definite descriptions as formally ungrammatical (ill-formed), free theories of definite descriptions hold all definite descriptions to be genuine singular terms, but do not regard unfulfilled definite descriptions as referring to existents if they refer at all (see **Descriptions**). More precisely, in free theories of definite descriptions, if the noun or noun phrase following the word ‘the’ in a definite description is not true of exactly one existent thing, then it refers to no existent thing if it refers at all.

Free theories of definite descriptions are all based on the basic principle that

Everything is such that *it* is the same as *the* if and only if *it* and *it only* is .

This principle is known as Lambert’s Law in the literature. Free theories of definite descriptions form a continuum. At one end there is a Russell-like theory which counts all atomic statements having at least one unfulfilled definite description as false, and contains scope distinctions *à la* Russell. At the other end there is a Frege-like theory which counts all atomic identity statements having only unfulfilled definite descriptions as true, and collapses all scope distinctions held to be dependent on definite descriptions. There are also a multitude of intermediate cases in the literature which, in contrast to the other traditions, is dramatic evidence of the fecundity of free logic in the logical treatment of definite descriptions.

In many, if not most, treatments of the logic of singular and general terms, supplemented by the logical modalities ‘it is necessary that’ and ‘it is possible that’, the purely quantificational fragment is free. For in such developments, the conditional

C If every (existent) thing exists, then Mars exists,

an instance of the classical principle of universal specification, is rejected on the ground that it leads, via unimpeachable modal principles, to the conditional

C * If necessarily every (existent) thing exists, then necessarily Mars exists.

But C * is false because though ‘Necessarily every (existent) thing exists’ is true, ‘Necessarily Mars exists’ is false. In contrast, the principle of restricted universal specification yields only the innocent and trivial

C + If necessarily every (existent) thing exists then necessarily Mars exists, *provided that* necessarily Mars exists.

The various treatments of free logic can influence the relationship between the meta-logical notion of reference and the logical notion of identity, the relationship reflected in the traditional adequacy condition that, for example, the singular term ‘the planet causing the perturbations in the orbit of Mercury’ refers to Vulcan if and only if the planet causing the perturbations in the orbit of Mercury is the same as Vulcan. To clarify the point, suppose the singular term ‘the planet causing the perturbations in the orbit of Mercury’ is replaced throughout the adequacy condition by the singular term ‘Vulcan’. Then one obtains the special case that ‘Vulcan’ refers to Vulcan if and only if Vulcan is the same as Vulcan. Clearly the traditional adequacy condition is sustained in negative free logics because both the reference claim and the identity claim are false since there exists no such object as Vulcan. Similarly, if the free logic is positive, but treats all singular terms as referring, in the spirit of Meinong, then the

traditional adequacy condition is also sustained because both the identity claim and the reference claim will be true. However, if the free logic is positive and it recognizes that some singular terms do not refer, in the spirit of Russell, the traditional adequacy condition may have to be altered; in such a version of free logic, it is false that ‘Vulcan’ refers to Vulcan if Vulcan does not exist, but usually it is true that Vulcan is identical with Vulcan. In such a treatment of positive free logic the relationship between reference and identity is expressed in the following way: ‘the planet causing the perturbations in the orbit of Mercury’ refers to Vulcan if and only if Vulcan exists and the planet causing the perturbations in the orbit of Mercury is the same as Vulcan.

Free logic also exerts an influence in set theory; in free set theory, for instance, the naïve axiom of set abstraction, from which Russell deduced his famous paradox, is assertible without condition (see [Paradoxes of set and property §4](#)). Because of the replacement of universal specification by the restricted version of that principle, the most that can be deduced from the statement ‘Everything is a member of the set of sets which are not members of themselves if and only if it is not a member of itself’, an instance of the naïve principle of set abstraction, is that the set of all sets which are not members of themselves does not exist.

A final application of free logic concerns the important topic of complex general terms. In certain developments of free logic, both positive and negative, devices exist for generating complex general terms out of ‘open’ sentences, which are expressions without truth-value of the form ‘it rotates’. For instance, the open sentence in question can be turned into the complex general term ‘object such that it rotates’ by prefixing to it the general term-forming operator ‘object such that’. Then applying that complex general term to the singular term ‘Vulcan’ yields an expression capable of having a truth-value, the statement ‘Vulcan is an object such that it rotates’. In some developments this statement is equivalent to the statement ‘Vulcan rotates’ if and only if it is true that Vulcan exists. The result is a free logic in which scope distinctions with respect to the connective ‘it is not the case that’ can be made over the whole class of singular terms and not just with respect to the class of definite descriptions. For instance, in the developments in question, ‘Vulcan is an object such that it does not rotate’ may be false, although ‘It is not the case that Vulcan is an object such that it rotates’ may be true. If, however, ‘Vulcan’ is replaced by ‘Mars’, the distinction between negation in the general term and negation over the whole statement collapses because the resulting statements have the same truth-value in virtue of the existence of Mars. Indeed, the means are available in these augmented free logics to make a distinction in general between *de re* and *de dicto* predicates or properties (see [De re/de dicto](#)). In such treatments, ‘*t* exists’ can now be defined even if identity is not present – as Arthur Prior once anticipated. The requisite definition is that ‘*t* exists’ means ‘For any general term *G*, it is *not* the case that *t* is an object such that it is *G* if and only if *t* is an object such that it is *not* the case that it is *G*’. More idiomatically, the definition says that an object exists just in case denying it is *G* is tantamount to asserting it is *non-G*, and vice versa, for any *G*.

See also: [Free logics](#); [Logical and mathematical terms, glossary of](#)

KAREL LAMBERT

References and further reading

- Bencivenga, E.** (1989) ‘Why Free Logic?’ in *Looser Ends*, Minneapolis, MN: University of Minnesota Press. (A very readable and distinctive philosophical explanation of free logic.)
- Evans, G.** (1979) ‘Reference and Contingency’, *The Monist* 62: 161-89. (An argument that Kripke’s doctrine of contingent a priori statements presupposes free logic.)
- Lambert, K.** (1983) *Meinong and the Principle of Independence*, Cambridge: Cambridge University Press, esp. ch. 5. (An informal account of the varieties of free logic and its connection with issues in the theory of predication.)
- Leonard, H.** (1956) ‘The Logic of Existence’, *Philosophical Studies* 7: 49-64. (The essay which initiated the development of free logics.)

Fuzzy logic

The term 'fuzzy' refers to concepts without precise borders. Membership in a 'fuzzy' set - the set of things to which a 'fuzzy' concept (fuzzily) applies - is to be thought of as being a matter of degree. Hence, in order to specify a fuzzy set, one must specify for every item in the universe the extent to which the item is a member of the set. The engineer Lotfi Zadeh developed a theory of fuzzy sets and advocated their use in many areas of engineering and science. Zadeh and his zealous followers have attempted to develop fuzzy systems theory, fuzzy algorithms and even fuzzy arithmetic. The phrase 'fuzzy logic' has come to be applied rather imprecisely to any analysis that is not strictly binary. It does not refer to any particular formal logic, in the sense in which the term 'logic' is used by philosophers and mathematicians. ('Fuzzy logic' is sometimes used anachronistically to refer to any many-valued logic.)

1 Philosophy

Fuzzy theory has, paradoxically, tended to partition the intellectual community into true believers and unbelievers. Early fuzzy researchers sometimes ignored or failed to credit properly ideas and similar approaches in other areas, resulting in much duplication of work. Simplistic condemnations of bivalent thinking and everything based on it (virtually all of science) were made, along with broad, often unsupported, claims for the utility of a fuzzy approach. These factors together with the intellectual conservatism of the engineering community led to the unwarranted view in some quarters that fuzzy theory was a fringe development designed only to appeal to those unable to measure up to the rigours of higher mathematical techniques. As a result, fuzzy theory was often ridiculed, frequently ignored, and sometimes excluded from conferences and publications. This situation resulted in some circling of the intellectual wagons and the founding of several conferences and journals devoted solely to fuzzy theory.

Some advocates have attributed the schism to philosophical differences, suggesting that a fuzzy world view has roots in eastern philosophies while western thinking has been dominated by bivalence. The fuzzy world view may be summed up by saying that everything is a matter of degree; that nothing real has any property to the complete exclusion of its negation. Fuzzy advocates criticize bivalent thinking as too simplistic.

Advocates of the fuzzy world view claim that bivalent thinking leads to paradoxes, such as the sorites paradox (see [Vagueness §2](#)). (If we pluck one hair from the face of a bearded man, he is still bearded; applying this argument many times over, there is no real difference between a bearded man and one without a beard.) They claim that we can avoid such paradoxes by recognizing that most predicates apply only to some degree. (If we pluck one hair from a man's face, then he is simply slightly less bearded than before: any sharp line drawn between having a beard and not having a beard is completely artificial.)

Critics respond that progress in science is usually the result of abstracting or idealizing: maps are useful because they leave out a lot of detail; computation of pressure/temperature/volume relationships does not require, and would be hindered by, the details of molecular positions, masses and velocities. Further, (western) science has developed mathematical techniques for dealing with matters of degree that do not require a fundamental change of logic. With regard to the paradoxes, critics point out that the specification of precise numerical degrees for concepts such as 'bearded' is artificial and impossible to carry out in any meaningful way. Further, there is no reason to think that we can linearly order the universe in terms of degree of 'beardedness'; some individuals are just not comparable. And if fuzzy advocates avoid drawing artificial distinctions, the paradoxes will simply reappear. For example, let 'really-bearded' apply 100% to those individuals whose degree of beardedness is 0.5 or greater and 0% to those whose degree of beardedness is less than 0.5. Unless fuzzy theorists are going to draw some very artificial fine distinctions, then it seems that removing one facial hair should not change one from being 'really-bearded'. If fuzzy advocates draw fine distinctions to avoid the paradox, then they can hardly criticize bivalent logic for drawing such distinctions.

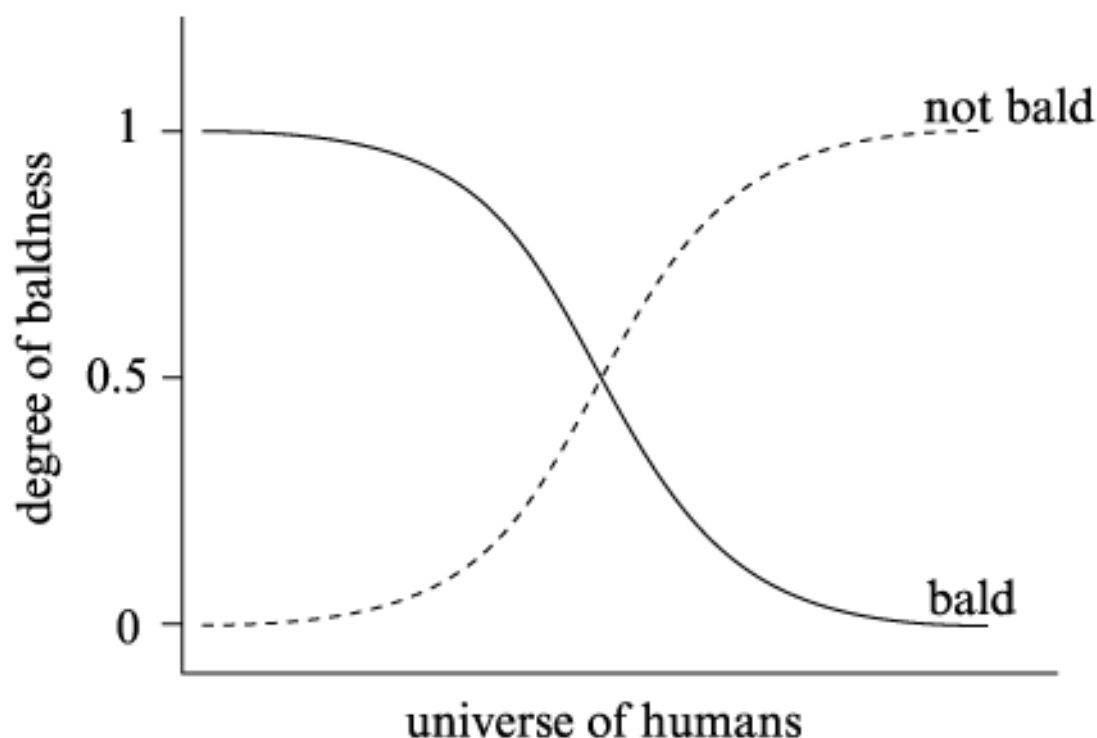
2 Vague concepts

Most of the concepts which we employ are vague rather than sharp. Vague concepts (see [Vagueness](#)) are extremely useful for getting around in the world and for communication. Suppose you have been told to look for a

tall, bald man with a beard, wearing a flowered shirt and walking with a limp. You do not need to have the criteria for the term 'tall' specified to within a millimetre; you do not need criteria for baldness spelled out in terms of percentage of coverage of scalp; and so on. If we tried to make these terms very precise, we would have more difficulty applying them. It seems that the more precise a term, the more difficult it is to determine whether it applies.

Classical logic makes the simplifying assumption of bivalence: every sentence is either true or false. But consider the concept of baldness. Certainly there are clear cases in which we would agree that 'X is bald' is true, and others in which we would agree that 'X is bald' is false. But there are also a lot of cases on the borderline. Instead of just 'true' or 'false', it is tempting to use locutions such as 'partially true', 'more true than not', and so on. Classical logic does not seem suited to reasoning with vague concepts. Such observations are cited as a motivation for many-valued logic (see [Many-valued logics](#)).

Max Black (1937) was the first to publish systematic suggestions for formalizing reasoning with vague concepts. Suppose our universe is human beings, and consider the predicate 'bald'. We may think of the predicate as standing for the set of objects which have the property. In classical logic we would have to decide all borderline cases arbitrarily. Classically, we may identify the set with its two-valued 'characteristic' function: $b(x) = 1$ if x is bald, and $b(x) = 0$ if x is not bald. Black suggested that instead of taking the characteristic function to be two-valued, we should allow it to take any value in the closed real interval $[0, 1]$. Ordering the universe in terms of degree of baldness, we could plot a 'baldness' curve as shown.



Black advocated treating the negation of a concept as the complement with respect to 1, that is, $\sim b(x) = 1 - b(x)$. He noted that the curve for a vague concept and its negation would cross at the 0.5 position. He further noted that for items in the grey area between 0 and 1, it made sense to say that the item both did and did not have the property. Black recognized the rather arbitrary nature of assigning precise numerical degrees to the membership relation. He suggested using statistical surveys and assigning $b(x)$ to be the percentage of the linguistic community that would agree that x has property b . Paradoxically, this procedure is based on the possibility of forcing vague concepts into a binary ideal that Black's analysis was intended to avoid.

3 Fuzzy theory

In name, fuzzy set theory was first developed by Lotfi Zadeh (1965, 1975), though the content is strikingly similar to Black's earlier analysis. However, Zadeh took the theory much further.

Assume some universe, U ; identify each fuzzy set f with its characteristic function $f : U \rightarrow [0, 1]$. Zadeh suggested the following definitions.

union: $(f \cup g)(x) = \max\{f(x), g(x)\}$
 intersection: $(f \cap g)(x) = \min\{f(x), g(x)\}$
 complement: $f'(x) = 1 - f(x)$
 universal set: $u(x) = 1$
 empty set: $\emptyset(x) = 0$

These definitions reduce to the classical notions if we restrict characteristic functions to being two-valued. Union and intersection so defined are commutative, associative, idempotent and mutually distributive, and De Morgan's laws are satisfied. (Zadeh's definitions are not the only ones to satisfy these properties.) However, if there is an x in U for which $0 < f(x) < 1$, then $f \cup f' \neq u$ and $f \cap f' \neq \emptyset$. These observations correspond well with our intuition that we cannot draw a sharp line between, for example, 'bald' and 'not bald'.

Difficulties arise when trying to treat fuzzy set theory as a basis for a logic of vagueness. There is a simple correspondence between classical set theory on a universal set and classical logic. We may think of the universe as the domain of a model, with sets corresponding to predicates; and union, intersection and complement corresponding to disjunction, conjunction and negation, respectively. The same sort of correspondence exists between Zadeh's fuzzy set theory and the infinite-valued logic of Łukasiewicz (§3). But the quantificational version of infinite-valued Łukasiewicz logic is not finitely axiomatizable. Further, trying to assign a precise numerical degree to the extent to which each person is bald seems even more artificial than simply assigning True or False to each attribution.

Finally, Łukasiewicz logic is not expressively strong enough to allow one to express in the object language the fundamental semantic concepts which served as the motivation. We want our formal language to contain 'autodescriptive' operators J_i for each truth-value i , such that, in any particular model, the sentence $J_i(E)$ takes the value 1 if E takes the value i in that model, and 0 otherwise. But if we include autodescriptive operators in the syntax, it is not possible to give a finitary proof theory for even the propositional logic. So, either the object language must be inadequate to express the notions fundamental to fuzzy theory, or there can be no finitary proof theory.

Zadeh's response to such criticisms was to fuzzify even the metatheory. He suggested that a fuzzy logic should be regarded as a complex association of some formal base logic (for example, that of Łukasiewicz) with fuzzy linguistic structures. Semantic attributions such as 'true', 'very true', 'somewhat true' and so on are to be treated as fuzzy sets. He claimed that all rules of inference are only approximate, criteria for good arguments can be given only semantically, and that the standard metatheoretical notions such as consistency and completeness are not relevant to fuzzy logic.

In spite of the difficulties of developing a formal logic based on fuzzy principles, a fuzzy analysis has led to some elegantly simple techniques for mathematical approximation of complex systems. As a result there have been some impressive systems developed in areas of expert systems and control engineering.

See also: [Logical and mathematical terms, glossary of](#)

CHARLES G. MORGAN

References and further reading

- Black, M.** (1937) 'Vagueness: An Exercise in Logical Analysis', *Philosophy of Science* 427-55. (Original analysis of vague concepts; referred to in §2.)
- Dubois, D. and Prade, H.** (1980) *Fuzzy Sets and Systems*, New York: Academic Press. (Excellent formal development and review of the literature of the theory of fuzzy sets and related matters.)
- Kosko, B.** (1992) *Neural Networks and Fuzzy Systems*, Englewood Cliffs, NJ: Prentice Hall. (Formal development of important results in fuzzy systems theory, along with partisan philosophical commentary on the fuzzy world view.)
- Kosko, B.** (1993) *Fuzzy Thinking*, New York: Hyperion. (A much less formal version of the previous item with some intellectual history.)

- Morgan, C.G. and Pelletier, F.J.** (1976) 'Some Notes concerning Fuzzy Logics', *Linguistics and Philosophy* 1: 79-97.(Problems finding proof theory for fuzzy logic.)
- Scarpellini, B.** (1962) 'Die Nicht-Axiomatisierbarkeit des unendlichwertigen Prädikatenkalküls von Łukasiewicz', *Journal of Symbolic Logic* 27: 159-70.(Quantified infinite-valued Łukasiewicz logic is not axiomatizable; see §3.)
- Zadeh, L.A.** (1965) 'Fuzzy Sets', *Information and Control* 8: 338-53.(Classic paper on fuzzy set theory; referred to in §3.)
- Zadeh, L.A.** (1975) 'Fuzzy Logic and Approximate Reasoning', *Synthese* 30: 407-28.(Rejection of formal proof theory for fuzzy logic; referred to in §3.)

Geometry, philosophical issues in

The least abstract form of mathematics, geometry has, from the earliest Hellenic times, been accorded a curious position straddling empirical and exact science. Its standing as an empirical and approximate science stems from the practical pursuits of land surveying and measuring, from the prominence of visual aids (figures and constructions) in geometric proofs and, in the twentieth century, from Einstein's General Theory of Relativity, which holds that the geometry of spacetime is dependent upon physical quantities. On the other hand, very early on, the symmetry and perfect regularity of certain geometric figures were taken as representative of a higher knowledge than that afforded by sense experience. And its concern with figures and constructions, rather than with number and calculation, rendered geometry amenable to axiomatic formulation and syllogistic deduction, establishing a paradigm of demonstrative knowledge which endured for two millennia. While the progress of mathematics has surmounted traditional distinctions between geometry and the mathematics of number, leaving only a heuristic role for geometric intuition, geometric thinking remains a vital component of mathematical cognition.

1 The idea of a demonstrative science

The earliest records of activity in geometry come from Babylon. What we know as Pythagoras' theorem appears in cuneiform texts of around 2600 BC, where it is given an empirical and approximate verification. As the name indicates, the subject of geometry originated in the practical pursuits of land measurement and surveying (though probably not first among the Egyptians, as Herodotus famously reports). Although the Greeks did not originate geometry, it is to them that we owe the conception of geometry as an exact demonstrative science as against an empirical and practical discipline. While anecdotal evidence identifies Thales of Miletus as the originator of the idea that geometric statements are to be proved, it is generally accepted that the Pythagoreans had transformed mathematics into a deductive science by around 500 BC. Pythagoras himself is said to have conferred upon geometry the standing of *theōria*, the manner of contemplative knowledge alone worthy of a free and not a slave people; and to have propounded the belief that its true subject matter pertained to intelligible objects rather than to sense experience. Later, in Plato, geometric figures - circles, triangles, the regular solids - became archetypal ideal forms: perfect, universal, absolute, eternal and harmonious; whereas their sensible instantiations in matter were seen as inexact, particular, relative, temporal and discordant (see [Pythagoreanism §2](#)).

Early Pythagorean methods of proof had probably not advanced much beyond primitive diagrammatic methods. By arranging pebbles in simple geometric arrays, a number of elementary theorems had been discovered by direct observation. Other visual methods included superposition, whereby the identity or congruence of two geometric figures was established by moving one figure into point-for-point coincidence with another. Some have conjectured that the first Greek proofs of the Pythagorean theorem were discovered in this manner. While such methods were suitable to the discovery of simple arithmetic and geometric facts, they were not as conclusive as the deductive methods canonically compiled in Euclid's *Elements of Geometry* (c.300 BC). Through its influence, the idea of a demonstrative science was established and geometry became the paradigm of systematic presentation of a body of knowledge in terms of logical deductions from axioms, whose truth was antecedently recognized.

At the basis of geometry, Euclid distinguished axioms, postulates and definitions. Euclid calls axioms 'common notions' which, as the starting points of demonstration, are held to be immediately evident, non-demonstrable truths or 'intelligible principles' deniable only on pain of absurdity. In this, he followed Aristotle rather than Plato, for whom a dialectical demonstration of the starting point of any particular systematic inquiry into the nature of things was required (*Republic VII.533c*). The five axioms all concern the concept of equality and might be considered as stipulating properties of identity applicable to magnitudes generally, for example, the third: 'Two things equal to a third thing are equal to each other.'

In addition to his axioms or 'common notions', Euclid also stated five postulates which are not asserted as true but rather as hypotheses to be tested through the conformity with experience of results deduced from them. The first three outline the permissible means by which geometric figures may be constructed: (1) by drawing a straight line from any point to any other point; (2) by producing a finite straight line continuously in a straight line; (3) by describing a circle with any given point as centre and any given distance as radius. These postulates essentially

delimit geometric construction as a procedure involving only straight edge and compass. This bounds what has been traditionally termed ‘elementary geometry’ and one may ask why such constructions were singled out, for at an early date Greek geometers had already recognized that there were many interesting problems in construction which did not appear to be solvable by these means alone. Three famous examples known in ancient times are: to construct a square having the area of a given circle (‘squaring the circle’); to trisect a given angle; and to construct a cube having double the volume of a given cube. [Plutarch \(c.46-120\)](#) reports that Plato was incensed by the attempts by Eudoxus and Archytas to solve the problem of doubling the cube by employing mechanical instruments in trial and error fashion; for Plato, such methods accorded sense experience an unwarranted legitimacy and hence were a corruption of the ‘pure excellence of geometry’. In fact, not until the nineteenth century were rigorous proofs given that these are impossible ruler and compass constructions, each additionally requiring some version of a principle of continuity.

Of the remaining two postulates, (4) All right angles are equal to each other, and (5) Given a line l and a point P not on it, there is a unique line parallel to l through P , (5) was often singled out as lacking the evidential force of the others. As a result, from Euclid’s time up to the early part of the nineteenth century, many attempts were made in vain to derive it from Euclid’s other assumptions. (Mathematical practice over the centuries had largely submerged Euclid’s evidential distinction between axioms and postulates.) Reflecting on this dismal record in 1819, C.F. Gauss had to admit that ‘In the theory of parallels, we are still no further than Euclid’, a situation he called ‘the shame of mathematics’. Resolution took quite an unexpected form. After unsuccessfully attempting to demonstrate the postulate, both J. Bolyai (in 1823) and N. Lobachevskii (in 1826) independently formulated a consistent non-Euclidean (hyperbolic) geometry in which the fifth postulate did not hold, the other Euclidean axioms and postulates remaining intact. Gauss himself had privately entertained such thoughts as early as 1799. Significantly, Gauss and Lobachevskii (but not Bolyai, who conceived of geometry as a purely logical system based upon axioms) seem to have thought that the truth as to the Euclidean or non-Euclidean character of geometry was an empirical matter to be resolved by certain observations in geodesy or astronomy (for example, stellar parallax). The mathematical and philosophical ramifications of these discoveries were slow to emerge; Bolyai and Lobachevskii had published obscurely and Gauss not at all. But by mid-century, a transformation was underway in the Euclidean conception of geometric knowledge as a unique body of truths concerning space whose certainty was secured by intuition.

2 Towards formal rigour

For nearly two millennia, Euclid’s *Elements* was the exemplar of exact reasoning. It is instructive to consider the developments which resulted in overturning this perspective. The first took place between the end of the seventeenth and the end of the eighteenth centuries. For Hobbes, Spinoza and Descartes, reasoning *more geometrico* (‘in the geometric manner’) attained the deductive ideal, whereby the truth of single propositions inexorably followed from propositions either antecedently established or accepted as fundamental truths. According to Pascal ([c.1657](#)), the ‘art of persuasion’ is identified with the method whereby evident axioms are laid down, clear definitions of non-primitive terms are given, and demonstration proceeds in step-by-step fashion. Even Newton, the co-discoverer of the differential and integral calculus, wrote his *Mathematical Principles of Natural Philosophy* ([1687](#)) in the axiomatic style of a classical geometric treatise, employing geometric methods of proof where the new methods of the calculus would have been more appropriate.

However, the enormous progress of mathematics in the eighteenth century was almost entirely due to developments in algebra and the new science of analysis, the study of functions and variables pioneered, above all, by Euler. Correspondingly, the guiding ideal of rigour in mathematical proof shifted from the figures and constructions of geometry to the permitted manipulations and transpositions of symbols characteristic of algebra and analysis. Thus Lagrange would famously boast in the beginning of his *Mécanique analytique* (1788) that ‘One will not find any figures in this work. The methods that I present require neither constructions, nor geometrical or mechanical reasoning, but only algebraic operations subjected to a regular and uniform development’. Another shift took place at the end of the nineteenth century in tandem with the programme of seeking greater rigour in the calculus, the so-called ‘arithmetization’ of analysis. Here, the analytic idea of rigour was carried into geometry itself in the demand that deduction not only be independent of diagrams and figures but also of the meaning of geometric concepts. In 1882, Pasch provided an initial account of Euclidean geometry as a hypothetico-deductive system in which the Euclidean definitions of ‘point’, ‘line’ and ‘plane’ were replaced by stipulated unanalysed

relations between these concepts. However, the classic exposition of this point of view is Hilbert's *Grundlagen der Geometrie* (1899) which famously posits 'three different systems of things...points...straight lines...and planes', entities defined only implicitly through relations designated by terms such as 'lie upon', 'between', 'parallel', 'congruent' and 'continuous' whose complete characterization, in turn, lies in those propositions derivable from the chosen axioms. In this abstract conception, there is no longer any trace of the traditional conception of geometry as a theory of *actual* space and its properties. By the early twentieth century, this trend towards axiomatic formalism had completely supplanted the Euclidean paradigm of an exact science.

3 The nature of geometry

According to a once-customary view of mathematics, the two primordial mathematical conceptions are the notions of whole number (as exhibited in counting or ordering) and continuous magnitude or mere extension (as determined in measurement); hence mathematics is ultimately rooted in the sciences of number (arithmetic, algebra, analysis) and the science of space (geometry). However, the conception of geometry as the science of space seems only to have emerged in the late Renaissance, as the modern notion of 'space' arose around 1600 AD (see *Space*), whereas the science of number has a pedigree reaching back to the speculations of the Pythagoreans. Indeed, the Pythagoreans initially did not clearly distinguish geometry from number. Number was the essential reality, though numbers appear to have been conceived as dots in particular geometric arrangements (thus, for example, the terms 'square number' and 'triangular number', which still survive). The Pythagorean discovery of incommensurable lines (for example, $\sqrt{2}$ which is the diagonal of a unit square) perhaps first effected a separation, for a 'horror of the irrational' prevented a generalization of the concept of number suitable for measurement as well as counting. Hence, irrational magnitudes could only be represented geometrically. As a result, the science of number remained paramount and, in the *Republic*, Plato places the study of geometry after the study of number on the grounds that, unlike the figures of geometry, numbers have no sensible and tangible bodies but are purely objects of thought accessible only to the soul (*VII.526*). A more compelling reason for the subordination is simply that geometry, in speaking of triangles, quadrilaterals and so on, presupposes number.

The ancient distinction of geometry and the science of number was blurred by the invention of analytic geometry by Fermat (1629) and Descartes (1637). The essential idea here - the introduction of numerical coordinates into the plane - establishes a one-to-one correspondence between plane curves and equations in two variables: to each distinct curve corresponds a definite equation $f(x, y) = 0$. This permits the study of the geometric properties of curves through algebraic or analytic relations connecting their equations, an innovation greatly expanding the variety of curves and figures which can be submitted to precise mathematical treatment. Moreover, coordinates provide a ready means of generalizing from two to three dimensions, and then, as Grassmann (1844) showed, to n -dimensional spaces. Thus it is only with the introduction of coordinates that geometry acquires the requisite generality of a 'science of space' (and indeed, more than the requisite generality); ironically, geometry obtains the status of 'science of space' through methods which enable a reduction of geometry to number.

However, the triumph of analytic methods in geometry was carefully circumscribed. Newton, for example, held that equations as 'Expressions of Arithmetical Computation...properly have no place in Geometry', and he celebrated 'the Ancients' over 'the Moderns', in that the former 'never introduced Arithmetical Terms into Geometry' whereas the latter 'by confounding both, have lost the simplicity in which all the Elegancy of Geometry consists' ([1707] 1728: 228; 1967: 120). Newton's intent, it seems, was to insulate only elementary geometry from algebraic methods but not higher geometry, where coordinate methods were actively employed and extended. Other dissenters from the 'triumph of analysis' soon emerged. Most influential were the early nineteenth century French school of projective geometers (L. Carnot, Chasles, Poncelet, among others). Expressing empiricist disdain for the use in calculus and algebra of such 'metaphysical' entities as infinitesimals and negative numbers, they championed instead coordinate-free 'synthetic' methods of proof whereby mathematical certainty was safeguarded by exclusive focus upon a constructed figure (see Poncelet 1822).

At issue here is the status of the diagrams employed in geometric proofs. Were constructed figures merely superfluous, *in concreto* visual aids to clarify bits of abstract reasoning in the text? Or were geometric truths somehow necessarily connected with the quasi-visual evidence provided by representation in 'the mind's eye' of intuition? For Kant, it was clearly the latter. Euclidean geometry characterizes that 'pure form of our sensible intuition' which is space, a condition imposed by the mind upon the experience of the outer sensible world and

thus a condition of the possibility of experience of this world. Furthermore, the ruler and compass constructions of Euclidean geometry are examples *par excellence* of how, for Kant, mathematical knowledge is attained solely through construction of concepts, that is, exhibition a priori of an intuition (for example, a figure such as a triangle) corresponding to its concept which, though particular, is none the less representative of all the universality contained in the concept. A geometric proof is thus ‘a chain of inference guided throughout by intuition’ (1781/1787: a717; b745).

Mathematical developments in the nineteenth century undermined Kant’s conception. First, as noted above, the discovery of consistent (that is, relative to Euclidean geometry) non-Euclidean geometries undercut the prevalent belief in the necessary truth of Euclidean geometry as a characterization of space. Second, the development of projective geometry led to a vastly generalized conception of a geometry as the study of properties of figures invariant under a given group of transformations. Thus, there are as many geometries as there are groups of transformations acting on a space. Third, under the influence of Riemann (see below), physical space came to be seen as a special three-dimensional case of a more general geometric structure. From each of these developments, it followed that the structure of physical space was no longer geometry’s sole subject matter. Thus different ‘mathematical spaces’ came to be distinguished from physical space, and both from a quasi-perceptual and psychological ‘space of intuition’. Finally, the trend towards greater rigour in mathematical proof in the latter part of the century denied intuition any intrinsic evidential standing. This in turn promoted axiomatic formalism in geometry, where an explicit recognition that geometric primitives have no pre-axiomatic meaning replaced the traditional view of its primitive terms as indefinable but intuitively secure concepts.

4 Geometry of physical space

Once the conceptual possibility opened that the geometry of physical space is not necessarily Euclidean, the question concerning the structure of this space appeared to be empirically resolvable through measurement. Poincaré (1902) argued, however, that any empirical determination of the geometry of physical space must rest on certain prior conventions concerning metric concepts, such as whether the path of a light ray is considered ‘straight’. Others saw a still more fundamental problem concerning what is presupposed by measurement at all. To this, Riemann (1854) and Helmholtz (1868) provided superficially similar but in fact quite different solutions. For both, measurement rests upon the possibility of free mobility of the standard of measurement. However, Riemann interpreted this to mean that magnitudes are independent of their position in what he termed an n -fold extended quantity, of which physical space is merely a particular three-dimensional instance (see [Space §4](#)). Riemann postulated an analytical expression, a generalization of the Pythagorean theorem to n dimensions, which is an invariant measure of the infinitesimal lengths between neighbouring points in such a structure, roughly the modern idea of a real n -dimensional differentiable manifold. Given this measure, one can define the arc length of a curve, the angle between intersecting curves, the volume of a region and other geometric concepts. This geometry is locally Euclidean in that it gives a Euclidean metric structure to the tangent space at each point P , hence the measure applies generally to flat and to curved manifolds and also to manifolds with variable (non-constant) curvature. However, in this latter type of manifold, Riemann went on to observe, the magnitude of a solid body is not independent of the position of the body.

Helmholtz, on the other hand, saw the postulate of free mobility as resting upon presumed facts about the behaviour of such bodies *in physical space*, in particular, that there are rigid motions. While acknowledging that the notion of a perfectly rigid body is an idealization to which nothing in the physical world corresponds, Helmholtz none the less saw this concept as a presupposition of measurement in general and hence of the geometry of physical space. As subsequently refined by Reichenbach (1928), who emphasized the impossibility of actually ascertaining whether a physical measuring rod undergoes alterations of length in transport, this view introduces an arbitrary element into measurement in that it makes determinations of the geometry of physical space rest on the adoption of conventions regarding the behaviour of rigid bodies. However, in the generic case of variably curved space-times permitted by Einstein’s Theory of General Relativity (1915), the notion of a rigid body is suspect for just the reason suggested 61 years earlier by Riemann: there are no congruences corresponding to the supposed invariant length of a rigid body. In this situation the most consistent procedure, as Weyl pointed out, is to renounce the concept of a transportable standard length. Weyl then showed (1921) how metric properties of the space-time manifold could be mathematically derived from weaker conformal and affine properties having, as their respective empirical correlates, two metric-independent physical processes, the paths of light rays (so-called null geodesics)

and the inertial trajectories of point masses (particles ‘too small’ to be affected by forces of surrounding physical fields). The essentials of this procedure remain a matter of live interest among general relativists (see Ehlers 1988).

5 But what is geometry?

With the triumph of formal axiomatics, geometry had lost its two most distinctive characteristics: its standing as the theory of physical space and the intuitional trappings attending its traditional reliance on figures and constructions. What then distinguishes geometry from any other formally axiomatized mathematical theory? Surveying the subject in 1932, two eminent geometers wrote:

The question then arises why the name geometry is given to some mathematical sciences and not to others. It is likely that there is no definite answer to this question, but that a branch of mathematics is called a geometry because the name seems good, on emotional and traditional grounds, to a sufficient number of competent people.

(Veblen and Whitehead 1932: 17)

Yet, as Cartan (1936) quickly pointed out, even in mathematics such a consensus may be difficult to achieve. This is not to say that boundaries are not still drawn. A leading contemporary geometer recently suggested that ‘a property is geometrical, if it does not deal directly with numbers or if it happens on a manifold, where the coordinates themselves have no meaning’ (Chern 1990: 685). In any case, it must be admitted that geometry has not been completely assimilated into the rest of mathematics, in part, perhaps, because the physical world continues to suggest geometric analogies. Moreover, mathematicians continue to speak of ‘geometric intuition’ as a fertile source of mathematical invention, even while recognizing that distinctions between geometry and other areas of mathematics remain a matter of taste. And although some mathematicians doubt that ‘the psychological aspects of true geometric intuition’ will ever be ‘cleared up’ (Weil 1978), others speculate that the visual-intuitive mode of cognition so characteristic of geometry stems from a lateral specialization of the brain localized in the right hemisphere of the cerebral cortex (Yaglom 1988: 26).

See also: *Conventionalism*; Einstein, A.; Helmholtz, H. von; Hilbert’s programme and formalism; Kant, I.; *Logical and mathematical terms, glossary of*; *Mathematics, foundations of*; Poincaré, J.H.; *Relativity theory, philosophical significance of*; *Spacetime*; Weyl, H.

T.A. RYCKMAN

References and further reading

- Atiyah, M.** (1982) ‘What is Geometry?’, *The Mathematical Gazette* 66: 179-84. (The suggestion, by an eminent geometer, that geometry is that part of mathematics where visual thought is dominant.)
- Boyer, C.** (1956) *History of Analytic Geometry*, New York: Scripta mathematica. (A comprehensive account of the innovations of Fermat and Descartes.)
- Cartan, E.** (1936) ‘Le rôle de la théorie des groupes de lie dans l’évolution de la géométrie moderne’, *Comptes rendus du congrès international des mathématiciens* 1: 92-103. (The case for continuous groups as the unifying concept in geometry.)
- Chern, S.-S.** (1990) ‘What is Geometry?’, *American Mathematical Monthly* 97: 679-86. (A survey from Euclid to fibre bundles.)
- Coolidge, J.** (1940) *A History of Geometrical Methods*, Oxford: Clarendon Press. (Compendium covering Babylonia to Weyl’s generalization of Riemannian geometry.)
- Daston, L.J.** (1986) ‘The Physicalist Tradition in Early Nineteenth Century French Geometry’, *Studies in the History and Philosophy of Science* 17: 269-95. (An informative essay on the geometric empiricism of an influential school of projective geometers.)
- Descartes, R.** (1637) *La Géométrie*, Leiden: Maire; trans. D.E. Smith and M.L. Latham, *The Geometry of René Descartes*, with a facsimile of the 1st edn, New York: Dover, 1954. (The first published account of analytic geometry.)
- Ehlers, J.** (1988) ‘Einführung der Raum-Zeit-Struktur mittels Lichtstrahlen und Teilchen’ (The Introduction of Space-Time Structure through Light Rays and Particles), in J. Audretsch and K. Mainzer (eds) *Philosophie und Physik der Raum-Zeit*, Mannheim: Wissenschaftsverlag, 145-62. (A discussion of the problem of measurement in General Relativity.)
- Euclid** (c.300 BC) *The Thirteen Books of Euclid’s Elements*, trans. T. Heath, Cambridge: Cambridge University

- Press, 2nd edn, 1956, 3 vols.(First articulated conception of geometry as a deductive science, based upon axioms, postulates and definitions.)
- Fermat, P. de** (1629) 'Ad Locos Planos et Solidos Isagoge', in *Oeuvres*, ed. P. Tannery and C. Henry, Paris: Gauthiers-Villars, 1891-1922, vol. 1, 91-110 (Latin); vol. 3, 85-101 (French); partial English translation in D.E. Smith (ed.) *A Source Book in Mathematics*, New York: Dover, 1984, vol. 2, 389- 96.(Fermat's discovery of analytic geometry, published posthumously.)
- Grassmann, H.** (1844) *Die lineale Ausdehnungslehre, ein neuer Zweig der Mathematik*, Leipzig: Otto Wigand, 2nd edn, 1862; trans. L.C. Kannenberg, 'The Linear Theory of Extension: A New Branch of Mathematics', in *A New Branch of Mathematics: The 'Ausdehnungslehre' of 1844, and Other Works*, La Salle, IL: Open Court, 1995.(Incipient general formulation of vector analysis in n -dimensional space.)
- Helmholtz, H. von** (1868) 'Über die Tatsachen, die der Geometrie zugrunde liegen', *Nachrichten der Königl. Gesellschaft der Wissenschaften zu Göttingen* 9: 193-221; trans. M. Lowe, 'On the Facts Underlying Geometry', in *Epistemological Writings*, ed. P. Hertz and M. Schlick, Dordrecht: Reidel, 1977, 39-71.(The view that measurement presupposes the notion of a rigid body.)
- Hilbert, D.** (1899) *Grundlagen der Geometrie*, Leipzig: Teubner, 7th edn, 1930; 2nd edn trans. L. Unger and P. Bernays, *Foundations of Geometry*, La Salle, IL: Open Court, 1971.(Influential abstract axiomatization of Euclidean geometry.)
- Kant, I.** (1781/1787) *Critique of Pure Reason*, trans. N. Kemp Smith, London: Macmillan, 1929.(The 'Transcendental Aesthetic' and 'Transcendental Doctrine of Method' sections give Kant's views on geometry.)
- Klein, F.** (1895) *Vorträge über ausgewählte Fragen der Elementargeometrie*, Leipzig: Barth; trans. *Famous Problems of Elementary Geometry*, New York: Dover, 1956.(Classic examination of ruler and compass constructions involving no higher mathematics.)
- Müller, I.** (1981) *Philosophy of Mathematics and Deductive Structure in Euclid's Elements*, Cambridge, MA: MIT Press.(A modern perspective on axiomatic method in Euclid.)
- Nagel, E.** (1939) 'The Formation of Modern Conceptions of Formal Logic in the Development of Geometry', *Osiris* 7: 142-224; repr. in *Teleology Revisited and Other Essays in the Philosophy and History of Science*, New York: Columbia University Press, 1982, 195-259.(Classic study of the rise of the 'formalistic' conception of geometry.)
- Newton, I.** (1687) *The Mathematical Principles of Natural Philosophy*, trans. A. Motte (1729) and ed. F. Cajori, Berkeley, CA: University of California Press, 1962, 2 vols.(Its propositions are demonstrated synthetically, 'that the system of the heavens might be founded on good geometry'.)
- Newton, I.** (1707) *Arithmetica Universalis*, Cambridge: Whiston; trans. Ralphson and Cunn (1722), *Universal Arithmetik: Or, A Treatise of Arithmetical Composition and Resolution*, London: Sener, 2nd very much revised edn, 1728; repr. in *The Mathematical Works of Isaac Newton*, ed. D.T. Whiteside, London and New York: Johnson Reprint, 1967, vol. 2, 1-257.(Criticism of the Cartesian association of algebra and geometry.)
- Pascal, B.** (c.1657) *De l'esprit géométrique*, fragments published posthumously in *Oeuvres complètes*, Paris: Desclée de Brouwer, 1991, vol. 3, 390-428; trans. R. Schofield in *Scientific Treatises*, Chicago, IL, and London: Encyclopaedia Britannica, 1952, 430-46.(The view that 'geometry has all the perfection men are capable of'.)
- Plato** (c.380-367 BC) *Republic*, trans. P. Shorey (1930), London and Cambridge, MA: Loeb Classical Library, Heinemann and Harvard University Press, 1969, 2 vols.(Book VII locates the study of geometry as knowledge of 'the eternally existent' in the curriculum for future guardians of the ideal state.)
- Plutarch** (c.46-120) *Lives*, vol. 5, Cambridge, MA: Harvard University Press (Loeb Classical Library), 1961, 472-3.(Report of Plato's wrath at empirical methods in geometry.)
- Poincaré, H.** (1902) *La Science et l'hypothèse*, Paris: Flammarion; trans. W. Greenstreet, *Science and Hypothesis*, New York: Dover, 1952, ch. 5.(Classic statement of geometric conventionalism.)
- Poncelet, J.V.** (1822) *Traité des propriétés projectives des figures (Treatise on the Projective Properties of Figures)*, Paris: Bachelier; trans. selections entitled 'On Projective Geometry', in D.E. Smith (ed.) *A Source Book in Mathematics*, New York: Dover, 1984, vol. 2, 315-23.(Sketch of the significance of coordinate-free projective methods.)
- Reichenbach, H.** (1928) *Philosophie der Raum-Zeit Lehre*, Berlin: de Gruyter; trans. M. Reichenbach and J. Freund, *The Philosophy of Space and Time*, New York: Dover, 1982.(The case for conventionalism in the context of the General Theory of Relativity.)
- Riemann, B.** (1854) 'Über die Hypothesen, welche der Geometrie zu Grunde liegen', *Abhandlungen der Königl. Gesellschaft der Wissenschaften zu Göttingen* 13: 133-52; trans. with mathematical commentary, 'On the Hypotheses which lie at the Foundations of Geometry', in M. Spivak, *A Comprehensive Introduction to*

- Differential Geometry*, Houston: Publish or Perish, 2nd edn, 1990, vol. 2, 135-53. (Densely compressed, the deep mathematical and philosophical content of this article repays study.)
- Szabó, A.** (1964) 'The Transformation of Mathematics into Deductive Science and the Beginnings of its Foundation on Definitions and Axioms', *Scripta Mathematica* 28: 27-48A, 113-39. (A scholarly yet speculative account of the Pythagorean origins of mathematics as a demonstrative science.)
- Tarski, A.** (1959) 'What is Elementary Geometry?', in L. Henkin, P. Suppes and A. Tarski (eds) *The Axiomatic Method*, Amsterdam: North Holland, 16-29. (A logical formulation of elementary geometry without a set-theoretic basis.)
- Torretti, R.** (1978) *Philosophy of Geometry from Riemann to Poincaré*, Dordrecht: Reidel. (A work of high mathematical and philosophical distinction.)
- Veblen, O. and Whitehead, J.H.C.** (1932) *The Foundations of Differential Geometry*, London: Cambridge University Press. (A concise account reflecting stimulus provided by the theory of General Relativity.)
- Weil, A.** (1978) 'S.S. Chern as Geometer and Friend', in S.-S. Chern, *Selected Papers*, New York and Berlin: Springer, 1989, ix-xii. (Assessment of the place of geometry within mathematics by a leading number theorist.)
- Weyl, H.** (1921) 'Zur Infinitesimal Geometrie: Einordnung der projektiven und konformen Auffassung', *Nachrichten der Königlichen Gesellschaft der Wissenschaften zu Göttingen, mathematisch-physikalische Klasse*, 99-112; repr. in *Gesammelte Abhandlungen*, Berlin and New York: Springer, 1968, vol. 2, 195-207. (The theorem that conformal and affine properties of a Riemannian manifold determine the metric.)
- Yaglom, I.M.** (1988) *Felix Klein and Sophus Lie; Evolution of the Idea of Symmetry in the Nineteenth Century*, trans. S. Sossinsky, Boston, MA, and Basle: Birkhäuser. (A readable account of the development of the group concept in geometry.)

Gödel's theorems

Utilizing the formalization of mathematics and logic found in Whitehead and Russell's *Principia Mathematica* (1910), Hilbert and Ackermann (1928) gave precise formulations of a variety of foundational and methodological problems, among them the so-called 'completeness problem' for formal axiomatic theories - the problem of whether all truths or laws pertaining to their subjects are provable within them. Applied to a proposed system for first-order quantificational logic, the completeness problem is the problem of whether all logically valid formulas are provable in it.

In his doctoral dissertation (1929), Gödel gave a positive solution to the completeness problem for a system of quantificational logic based on the work of Whitehead and Russell. This is the first of the three theorems that we here refer to as 'Gödel's theorems'.

The other two theorems arose from Gödel's continued investigation of the completeness problem for more comprehensive formal systems - including, especially, systems comprehensive enough to encompass all known methods of mathematical proof. Here, however, the question was not whether all logically valid formulas are provable (they are), but whether all formulas true in the intended interpretations of the systems are.

For this to be the case, the systems would have to prove either S or the denial of S for each sentence S of their languages. In his first incompleteness theorem, Gödel showed that the systems investigated were not complete in this sense. Indeed, there are even sentences of a simple arithmetic type that the systems can neither prove nor refute, provided they are consistent. So even the class of simple arithmetic truths is not formally axiomatizable.

The idea behind Gödel's proof is basically as follows. Let a given system T satisfy the following conditions: (1) it is powerful enough to prove of each sentence in its language that if it proves it, then it proves that it proves it, and (2) it is capable of proving of a certain sentence G (Gödel's self-referential sentence) that it is equivalent to ' G is not provable in T '. Under these conditions, T cannot prove G , so long as T is consistent. For suppose T proved G . By (1) it would also prove ' G is provable in T ', and by (2) it would prove ' G is not provable in T '. Hence, T would be inconsistent.

Under slightly stronger conditions - specifically, (2) and (1') every sentence of the form ' X is provable in T ' that T proves is true - it can be shown that a consistent T cannot prove ' $\text{not } G$ ' either. For if ' $\text{not } G$ ' were provable in T it would follow by (2) that ' G is provable in T ' would also be provable in T . But then by (1') G would be provable. Hence, T would be inconsistent.

The proof of Gödel's second incompleteness theorem essentially involves formalizing in T a proof of a formula expressing the proposition that if T is consistent, then G . The second incompleteness theorem (that is, the claim that if T is consistent it cannot prove its own consistency) then follows from this and the first part of the proof of the first incompleteness theorem.

The two incompleteness theorems have been applied to a wide variety of concerns in philosophy. The best known of these are critical applications to Hilbert's programme and logicism in the philosophy of mathematics and to mechanism in the philosophy of mind.

1 Background

Interest in questions of completeness dates from quite early on in the history of logic and the foundations of mathematics. Aristotle (*Posterior Analytics* 87b) described the goal of science as the complete identification of the laws pertaining to its various areas. Much later, Kant expressed concern with completeness when he described human reason as inevitably disposed to pose questions to itself that it cannot answer. Clearer still, perhaps, are various statements in letters by Gauss (1817, 1829), in which he asserted that it may in principle be impossible for us to give a complete specification of the laws of geometry a priori owing to the fact that its object, space, is at least partially 'external' to our minds. (See Kronecker ([1887] 1932: 265) for an expression of the same view.) Finally, we find something approaching modern formulations of the notion in Hilbert ([1900] 1930: 242), where it is said to be a duty to show that a set of axioms for geometry capture every geometric law, and in Huntington (1905: 17), where a set of postulates for real algebra is said to form a 'complete' axiomatic basis for the subject.

However, it was not until the appearance of Hilbert and Ackermann's *Grundzüge der theoretischen Logik* (1928: 42-3) that the completeness problem received its first fully explicit, precise and general formulation. Hilbert and Ackermann believed that the formal-symbolic approach of modern mathematics had contributed greatly to its remarkable development, and they therefore advocated taking the same approach to logic. Such an approach would, they believed, pave the way for a mathematically precise study of both logical and mathematical thinking, and would make it possible to settle in a mathematical way various of the methodological and foundational questions pertaining to them. Among these was the question of 'completeness' - that is, whether the full class of truths or laws properly pertaining to a given domain of thought can be proved in a given system purporting to formalize it. Whitehead and Russell (1910) presented a formalism (PM) for logic and mathematics, and this became the focus of the early investigations of the completeness problem.

The first results were obtained (independently) by Bernays, in his unpublished *Habilitationsschrift* of 1918 (see Bernays 1926), and Post, in his doctoral dissertation of 1920 (see Post 1921). They isolated the propositional part of Whitehead and Russell's PM and showed that it proved all valid sentences of the propositional variety. Post indeed proved the stronger result (known as 'Post-completeness') that no sentence of the propositional variety can be added to PM without creating an inconsistency.

Hilbert and Ackermann observed that the broader logical system known as the 'restricted', or 'first-order' predicate calculus is not Post-complete (1928: 92). With this as his point of departure, Gödel, in his doctoral dissertation (1929), then took up the ordinary completeness problem for this system. Using ideas introduced by Löwenheim (1915) and Skolem (1923), he obtained the basic positive result that every valid formula of the first-order quantificational variety is provable in PM. Indeed, he proved the somewhat stronger result that every first-order quantificational formula is either refutable in PM or is satisfiable in a denumerable domain. He then extended the basic result to first-order logic with identity and the stronger result from single formulas to denumerable systems of formulas. In this latter extension, he made use of a 'compactness' result (namely, that a denumerably infinite set of formulas is satisfiable if and only if each of its finite subsets is) that he proved by semantic means.

In the introduction to his dissertation (deleted from the version published in 1930), Gödel expressed awareness of the possibility that the completeness problem for other areas of mathematical thought might not result in the same positive type of result as the completeness problem for PM. It was therefore not surprising that in 1930 he turned his attention to the completeness problem for the architectonic systems of Whitehead and Russell and Zermelo and Fraenkel (see Gödel 1930, 1931). These systems interested Gödel precisely because of their comprehensiveness. Such comprehensiveness might, he noted, lead some to conjecture that they can decide all mathematical questions formulable in their languages. What Gödel showed was that this is mistaken. There are problems - indeed, relatively elementary problems of the arithmetic of the integers - that even the most comprehensive systems cannot decide. Gödel's investigation of the completeness problem thus ended with a negative result.

There is a relationship between Gödel's 1929/1930 and 1931 studies that seems worth noting, even though he himself did not note it. It stems from the fact that every formal system is composed of both a logical and a non-logical part, deficiencies in either of which can lead to incompleteness. A defect of the first sort would mean that though the non-logical part might itself be perfect in content, the logical part would not be able fully to extract that content.

Gödel ([1929] 1930) can be seen as showing that incompleteness in systems whose logic is rightly taken to be first-order is due to deficiencies in the non-logical part, and his 1931 paper can then be seen as showing that such deficiencies extend to a wide range of systems. The latter also points to *logical* deficiencies in another class of systems, however - namely, systems whose non-logical parts are composed of the second-order Peano postulates and whose logical part is some formalizable scheme of second-order logical principles. This is a result of the categoricity of the second-order Peano postulates (first established in Dedekind 1888). Being categorical, they are in a certain sense perfect in their content. Still, Gödel (1931) establishes that any formal system based on them must be incomplete. Hence, any formalizable scheme of second-order logical consequence must fail to capture some of the content latent in them. In other words, any formal system of second-order logic must fail to capture some valid second-order logical principle. The 1931 paper can thus be seen as rounding out the logical investigations of the 1930 paper by providing a negative solution to the completeness problem for second-order

logic.

2 The completeness theorem

Gödel's original completeness proof was based on a normalization of quantificational formulas developed by Löwenheim (1915) and Skolem (1923). According to this normalization, there is an effective way of finding, for each formula ϕ of a first-order quantificational language L , a formula $\sigma(\phi)$ such that ϕ is satisfiable iff $\sigma(\phi)$ is, and $\sigma(\phi)$ is of the form $U-E-QF$, where U is a (non-empty) string of universal quantifiers, E a (possibly empty) string of existential quantifiers and QF a quantifier-free formula. Gödel showed how to construct conjunctions $K_n = QF_1 \& \dots \& QF_n$ of instances QF_i of QF such that either for every n , K_n is truth-functionally satisfiable, or, for some n , K_n is not. Like Skolem, he then showed that if the first alternative holds, $\sigma(\phi)$ is satisfiable in the domain of the natural numbers. Finally, he was able to show - and this was his distinctive contribution - that if the second alternative holds, $\sigma(\phi)$ is refutable (that is, $\vdash_{PM} \neg\sigma(\phi)$). With this came the crucial link between satisfiability and derivability in PM needed to prove PM's completeness.

It follows that $\sigma(\phi)$ is either denumerably satisfiable or refutable, and from this and the fact (also used by Gödel) that $\sigma(\phi)$ is refutable only if ϕ is, it then follows that for any formula ϕ of the language of the pure predicate calculus either ϕ is (denumerably) satisfiable or ϕ is refutable.

Gödel's proof was given for a specific logical calculus having a countable language (that is, a language containing countably many individual constants). Since then, completeness proofs have been devised both for a wider variety of calculi having countable languages (see, for example, Henkin 1949; Hasenjäger 1953; Hintikka 1955) as well as for calculi with uncountable languages (see Maltsev 1936). From this work, the following general method for constructing models for consistent sets of sentences of infinite cardinality has emerged.

To construct a model of infinite cardinality α for a consistent set S of sentences of language L of cardinality not exceeding α , one begins by adding α -many constants to L to form the language L^* . One then extends S to a maximal consistent and term-complete set S^* . (A consistent set is 'maximal consistent' iff no set properly containing it is consistent. A set is 'term-complete' just in case for every formula ϕ and all terms $t_1, \dots, t_i, \dots, t_\alpha$ of L^* , if $\phi(t_1), \dots, \phi(t_i), \dots, \phi(t_\alpha)$ are all elements of the set, then so is $\forall x\phi(x)$.) Next one constructs a model for S^* whose domain is the class of constants of L^* . And finally, one restricts this model to L and automatically obtains a model for S .

The proof that this construction yields a model M proceeds roughly as follows. (1) For every closed term c that appears in S^* , let M assign c itself as value. In this way, every element of the domain of M becomes the referent of a term of L^* . (2) For each basic n -ary predicate P^n of L^* , let M assign the n -tuple $\langle t_1, \dots, t_n \rangle$ to the extension of P^n iff $P^n(t_1, \dots, t_n)$ is an element of S^* . In this way, all atomic sentences and all denials of atomic sentences that are elements of S^* become true in M . This is so because S^* , being consistent, does not contain any atomic sentence that is denied by another element of S^* .

This completes the basis clause. In the induction step one calls upon the fact that, owing to its maximal consistency (and, in the case of quantified sentences, its term-completeness), S^* contains, for each of its compound elements σ_c (of complexity greater than that of the denial of an atomic sentence), a set of less complex elements whose joint truth suffices for the truth of σ_c . These less complex sentences fall under the hypothesis of the induction. The truth of the sentences of the basis clause thus 'induces' the truth of all elements of S^* of the next higher degree of complexity, and this process continues through all degrees of complexity. The truth-in- M of the sentences covered in the basis step thus eventually 'oozes up' to each element of S^* .

Every consistent set of sentences thus has a model. From this, both 'weak' completeness (every valid sentence is derivable) and 'strong' completeness (every sentence validly implied by a set of sentences is derivable from it) follow for systems admitting the usual types of derivations.

3 The first incompleteness theorem

Gödel made use of a formalization of mathematics in his 1931 paper that was related to that given in Whitehead and Russell (1910). But he added a new device - the 'arithmetization' of formal systems. This was based upon a division of the objects of a formal language into three basic categories: (1) the basic symbols of the language; (2)

finite sequences of these symbols; and (3) finite sequences of finite sequences of them. In order to express properties and propositions concerning these objects as predicates and sentences of arithmetic, Gödel defined a correlation (now called a ‘Gödel-numbering’) between the objects of these three types and the natural numbers. He coded the individual symbols of the systems by distinct odd numbers. To each finite sequence $\sigma_1, \dots, \sigma_n$ of symbols he then assigned the number

$$2^{g(\sigma_1)} \cdot 3^{g(\sigma_2)} \cdot \dots \cdot p_n^{g(\sigma_n)},$$

where p_n is the n th prime number and $g(\sigma_i)$ ($1 \leq i \leq n$) the Gödel number of σ_i . Finally, to each finite sequence of finite sequences of symbols $\Sigma_1, \dots, \Sigma_k$, he assigned the number

$$2^{g(\Sigma_1)} \cdot 3^{g(\Sigma_2)} \cdot \dots \cdot p_k^{g(\Sigma_k)},$$

where p_k is the k th prime number and $g(\Sigma_i)$ ($1 \leq i \leq k$) the Gödel number of Σ_i .

Infinitely many others schemes would be possible. The essential features are that distinct expressions are assigned distinct numbers and that there are effective procedures for encoding expressions and for decoding numbers to find which expression (if any) they represent.

Under such an encoding, certain formulas express not only arithmetic statements about numbers but also metamathematical statements about such objects as formulas, axioms and proofs. Using this observation, Gödel exploited an analogy with the Richard and liar paradoxes (see [Paradoxes of set and property §6](#)). (In passing he also observed that any other ‘epistemological antinomy’ could be put to similar use. See Gödel (1931: 598).) The analogy runs as follows.

Suppose that a theory T proves only true sentences and that a sentence G ‘says’ of itself that it is not provable in T . Neither G nor $\neg G$ is then provable in T . For if G is true then it is not provable in T , and if G is false then it cannot be provable because T proves only true sentences. So G is not provable in T and hence true. Its denial, $\neg G$, is thus false and hence unprovable in T since T proves only true sentences.

Gödel’s proof follows this reasoning with one exception that he himself emphasized (1931: 599): it replaces the assumption that T proves only true sentences with the weaker assumption of ω -consistency (see below). The core of the proof is the construction of the ‘self-referential’ sentence G in the so-called ‘diagonal lemma’. In its general form, this lemma states that for every formula F of L_T (that is, the language of T) of one free variable, there is a sentence D of L_T such that $\vdash_T D \leftrightarrow F(\ulcorner D \urcorner)$ (the term ‘ $\ulcorner D \urcorner$ ’ stands for the numeral of L_T whose referent under the standard interpretation of L_T is the Gödel number of D). We need only the following specialization of the diagonal lemma to G :

(DL) For every theory T that contains a certain fragment of arithmetic, there is a sentence G of L_T such that $\vdash_T G \leftrightarrow \neg \text{Prov}_T(\ulcorner G \urcorner)$.

This lemma is also referred to as Gödel’s ‘fixed point theorem’, A being a ‘fixed point’ in T of a formula $\phi(x)$ just in case $\vdash_T A \leftrightarrow \phi(\ulcorner A \urcorner)$. The formula $\text{Prov}_T(x)$ ‘expresses’ the notion of a formula’s being provable in T in the limited sense that the set of A such that $\vdash_T \text{Prov}_T(\ulcorner A \urcorner)$ is the set of theorems of T . In the standard terminology, $\text{Prov}_T(x)$ ‘weakly represents’ the set of Gödel numbers of theorems of T in T ; that is

(WR) For every sentence A of L_T , $\vdash_T A$ iff $\vdash_T \text{Prov}_T(\ulcorner A \urcorner)$.

From DL and WR, the proof of Gödel’s first incompleteness theorem (G1) proceeds in two parts.

Part 1. If T is a consistent theory and $\text{Prov}_T(x)$ satisfies both WR in the ‘only if’ direction and DL, then $\not\vdash_T G$.

Proof. Suppose that

(1) $\vdash_T G$.

From this and WR in the ‘only if’ direction, we get

(2) $\vdash_T \text{Prov}_T(\ulcorner G \urcorner)$.

But by (1), DL and the assumption that the logic of T allows *modus ponens* inference, we get

$$\vdash_T \neg \text{Prov}_T(\ulcorner G \urcorner),$$

which together with (2) implies the inconsistency of T .

Part 2. If T is consistent and $\text{Prov}_T(x)$ satisfies both WR in the ‘if’ direction and DL, then $\nvdash_T \neg G$.

Proof. Suppose that

$$(3) \vdash_T \neg G.$$

From (3), DL and the fact that the logic of T allows both *modus tollens* inference and inference from $\neg\neg A$ to A , it follows that

$$\vdash_T \text{Prov}_T(\ulcorner G \urcorner).$$

But this and WR in the ‘if’ direction imply that

$$\vdash_T G,$$

which together with (3) implies the inconsistency of T .

Together, parts 1 and 2 establish

(G1) If T is consistent and $\text{Prov}_T(x)$ satisfies DL and WR, then $\nvdash_T G$ and $\nvdash_T \neg G$.

It should be noted that WR in the ‘if’ direction is a different condition from that of the consistency of T . It forces all theorems of T of the form $\text{Prov}_T(\ulcorner A \urcorner)$ to be true. Equivalently, given the decidability in T of the arithmetized proof relation of T , it forces T to be ω -consistent (that is, it forces T not to prove, for any formula ϕ , both all instances of $\phi(n)$ and also $\exists x \neg \phi(x)$). ω -consistency, which was explicitly used as a condition in Gödel’s original proof, thus figures in our proof as well. Using a more complicated fixed point formula than G and a more complicated proof than the one just given, Rosser (1936) proved a version of G1 that does not require ω -consistency.

Parts 1 and 2 reveal the basic structure of the proof for G1. The details that remain are proofs of DL and WR. These can be found in many standard logic texts (for example, Enderton 1972; Boolos and Jeffrey 1974). Since Gödel’s proof in 1931, a mathematically more interesting case of a purely combinatorial sentence undecidable in first-order Peano arithmetic (PA) has been given, in Paris and Harrington (1977). In the early 1980s, Friedman gave a still more striking case (involving Kruskal’s theorem) of an arithmetical sentence undecidable in fragments of second-order Peano arithmetic considerably stronger than PA.

4 The second incompleteness theorem

Gödel arrived at the idea for his second incompleteness theorem (G2) by noting that, using the provability expression $\text{Prov}_T(x)$ of part 1 of the proof of G1, one can define a consistency formula Con_T for T and that the proof of part 1 can then be carried out in T . In other words, one can prove ‘ $\text{Con}_T \rightarrow G$ ’ in T . Together with part 1, this then implies the unprovability in T of Con_T .

The proof of ‘ $\text{Con}_T \rightarrow G$ ’, however, requires conditions on $\text{Prov}_T(x)$ that are different from those (namely DL and WR) required for the proof of G1. What exactly these conditions are becomes clear from the proof of ‘ $\text{Con}_T \rightarrow G$ ’ given below.

The details of this proof require that we choose a specific definition of the consistency expression, Con_T . The definition we choose is the basic one, the formula ‘ $\forall x(\text{Prov}_T(x) \rightarrow \neg \text{Prov}_T(\text{neg}(x)))$ ’ (where $\text{neg}(x)$ stands for the function that sends the Gödel number of a formula to the Gödel number of its negation). We could define Con_T differently without altering the conditions on $\text{Prov}_T(x)$ that are required for the proof of ‘ $\text{Con}_T \rightarrow G$ ’ in T . The present definition, however, has the virtues of being both familiar and basic. We therefore proceed with the proof of

$\vdash_T \text{Con}_T \rightarrow G$. (G2 lemma)

Its core element is the construction of a derivation in T of $\neg \text{Con}_T$ from $\neg G$. We begin by stating five schemata whose provability in T will be assumed in the proof:

- (a) $\text{Prov}_T(\ulcorner A \urcorner) \rightarrow \text{Prov}_T(\ulcorner \text{Prov}_T(\ulcorner A \urcorner) \urcorner)$
- (b) $\text{Prov}_T(\ulcorner A \leftrightarrow B \urcorner) \rightarrow \text{Prov}_T(\ulcorner \neg B \leftrightarrow \neg A \urcorner)$
- (c) $\text{Prov}_T(\ulcorner A \leftrightarrow B \urcorner) \rightarrow \text{Prov}_T(\ulcorner A \rightarrow B \urcorner)$
- (d) $\text{Prov}_T(\ulcorner A \rightarrow B \urcorner) \rightarrow (\text{Prov}_T(\ulcorner A \urcorner) \rightarrow \text{Prov}_T(\ulcorner B \urcorner))$
- (e) $\text{Prov}_T(\ulcorner A \urcorner) \rightarrow \text{Prov}_T(\ulcorner \neg \neg A \urcorner)$.

The proof proper begins by taking $\neg G$ as a hypothesis in T . We then reason in T as follows. Applying DL (whose content ' $G \leftrightarrow \neg \text{Prov}_T(\ulcorner G \urcorner)$ ' is available in T), we get

(4) $\text{Prov}_T(\ulcorner G \urcorner)$.

From (4) and an instance of (a), we get

(5) $\text{Prov}_T(\ulcorner \text{Prov}_T(\ulcorner G \urcorner) \urcorner)$.

We now add a T -theoretic 'formalization' of DL; that is,

$\text{Prov}_T(\ulcorner G \leftrightarrow \neg \text{Prov}_T(\ulcorner G \urcorner) \urcorner)$.

This together with an instance of schema (b) gives

$\text{Prov}_T(\ulcorner \neg \neg \text{Prov}_T(\ulcorner G \urcorner) \leftrightarrow \neg G \urcorner)$,

which together with an instance of (c) then yields

$\text{Prov}_T(\ulcorner \neg \neg \text{Prov}_T(\ulcorner G \urcorner) \rightarrow \neg G \urcorner)$.

From this and an instance of (d) we get

$\text{Prov}_T(\ulcorner \neg \neg \text{Prov}_T(\ulcorner G \urcorner) \urcorner) \rightarrow \text{Prov}_T(\ulcorner \neg G \urcorner)$,

which with an instance of (e) gives us

$\text{Prov}_T(\ulcorner \text{Prov}_T(\ulcorner G \urcorner) \urcorner) \rightarrow \text{Prov}_T(\ulcorner \neg G \urcorner)$.

This and (5) via *modus ponens* in T imply

(6) $\text{Prov}_T(\ulcorner \neg G \urcorner)$.

(6) together with (4) then gives us

$\neg(\text{Prov}_T(\ulcorner G \urcorner) \rightarrow \neg \text{Prov}_T(\ulcorner \neg G \urcorner))$.

This and the definition of Con_T imply $\neg \text{Con}_T$. This is what was to be proved in T from our hypothesis $\neg G$. Discharging this hypothesis, we then get

$\neg G \rightarrow \neg \text{Con}_T$,

and from this, by contraposition in T , we conclude

$\text{Con}_T \rightarrow G$.

Inspecting this proof we see that it makes various common assumptions concerning the logic of T ; for example, that it allows *modus ponens* inference, classical contraposition and universal instantiation. It also utilizes various conditions on $\text{Prov}_T(x)$ - specifically, schemata (a)-(e) and the 'formalized' version of DL. All of these, we will now show, can be obtained from the following 'derivability conditions'.

- (DC1) For every sentence A of L_T , $\vdash_T A$ only if $\vdash_T \text{Prov}_T(\ulcorner A \urcorner)$.
 (DC2) For all formulas A and B of L_T ,
 $\vdash_T \text{Prov}_T(\ulcorner A \rightarrow B \urcorner) \rightarrow (\text{Prov}_T(\ulcorner A \urcorner) \rightarrow \text{Prov}_T(\ulcorner B \urcorner))$.
 (DC3) For every sentence A of L_T , $\vdash_T \text{Prov}_T(\ulcorner A \urcorner) \rightarrow \text{Prov}_T(\ulcorner \text{Prov}_T(\ulcorner A \urcorner) \urcorner)$.

The sufficiency of DC1-3 can be seen as follows. Schema (a) is simply DC3 and schema (d) is DC2. The ‘formalized’ version of DL used above is obtained immediately by applying DC1 to DL. Schema (b) is obtained by applying DC1 to the logical truths of the form $(A \leftrightarrow B) \rightarrow (\neg B \leftrightarrow \neg A)$ to get the several instances of $\text{Prov}_T(\ulcorner (A \leftrightarrow B) \rightarrow (\neg B \leftrightarrow \neg A) \urcorner)$, and then applying DC2. The remaining schemata are similarly obtained. Schema (c) is obtained by applying DC1 to the logical principle $(A \leftrightarrow B) \rightarrow (A \rightarrow B)$ to get $\text{Prov}_T(\ulcorner (A \leftrightarrow B) \rightarrow (A \rightarrow B) \urcorner)$, which then, by DC2, yields (c). Schema (e) is obtained by application of DC1 to $A \rightarrow \neg \neg A$ to get $\text{Prov}_T(\ulcorner A \rightarrow \neg \neg A \urcorner)$, and then application of DC2.

We thus arrive at the following generalized version of Gödel’s second incompleteness theorem.

(G2) If T is consistent, Con_T is defined as $\forall x(\text{Prov}_T(x) \rightarrow \neg \text{Prov}_T(\text{neg}(x)))$ and $\text{Prov}_T(x)$ satisfies DC1-3, then $\not\vdash_T \text{Con}_T$.

5 The derivability conditions

DC1-3 represent a streamlining of the original ‘Ableitbarkeitsforderungen’ used in Hilbert and Bernays (1939: 285-6) to give the first full proof of G2. The improvements are due mainly to Löb (1955), though his aim was to identify a set of conditions on $\text{Prov}_T(x)$ sufficient to solve a problem of Leon Henkin’s and not to provide a general set of conditions under which G2 holds.

Jeroslow (1973) produced a further streamlining of the derivability conditions by showing how to eliminate DC2. The key idea was a new fixed-point or diagonalization construction - referred to as a ‘literal’ fixed point. It states that

For every formula $\psi(x)$ of L_T of one free variable, one can effectively find a formula ϕ and a closed term τ of L_T such that ϕ is $\psi(\tau)$, and $\vdash_T \tau = \ulcorner \phi \urcorner$.

What Jeroslow proved was a version of G2 lemma with a literal Gödel sentence G_{Lit} - that is, a sentence of the form $\neg \text{Prov}_T(\tau)$, where $\vdash_T \tau = \ulcorner G_{\text{Lit}} \urcorner$, and not merely a sentence provably equivalent in T to the sentence $\neg \text{Prov}_T(\ulcorner G_{\text{Lit}} \urcorner)$ - in place of the usual Gödel sentence G . His proof relies only upon the literal diagonal lemma, DC1 and DC3. Likewise, the relevant part of Jeroslow’s version of G1 - namely, that if T is consistent, then $\not\vdash_T G_{\text{Lit}}$ - relies only upon the literal diagonal lemma and DC1. Together these two results yield a version of G2 that requires only the literal diagonal lemma, DC1 and DC3. We thus see that G2 can hold even when the formula expressing provability does not satisfy DC2. Among other things, this appears to distinguish G2 from Löb’s theorem (that is, the theorem that for any sentence A of L_T and any formula $\text{Prov}_T(x)$ satisfying DC1-3, if $\vdash_T \text{Prov}_T(\ulcorner A \urcorner) \rightarrow A$, then $\vdash_T A$), since the known proofs of the latter demand satisfaction of DC2 (see Kreisel and Takeuti 1974: 7-8).

The approaches of Hilbert and Bernays (1939), Löb (1955) and Jeroslow (1973) are what might be called ‘derivability conditions’ approaches since they all rely for their application upon the satisfaction of some set of abstract conditions on $\text{Prov}_T(x)$. Feferman (1957, 1960) has criticized such approaches on the ground that, for a given theory T and a given formula $\text{Prov}_T(x)$ weakly representing the theorems of T , it is often difficult to determine whether or not such a set of conditions holds. He therefore sought a condition on provability expressions whose satisfaction or non-satisfaction is easier to determine, but which is still sufficient for G2.

By requiring that provability expressions preserve the logical and inductive structures of the usual definitions of formal provability, Feferman was able to reduce his condition to one governing the choice of expressions for the axioms of a system. That requirement is that the axiomhood formula be RE, where, roughly, this means that it preserves the way in which the axioms are ‘presented’. An RE (‘recursively enumerable’) formula is one which transparently defines a recursively enumerable set of axioms. An RE formula representing an infinite set of axioms (which is the difficult case) imitates the inductive character of the metamathematical definition of that set (see

Smullyan 1961; Kreisel 1965; Feferman 1982, 1989). The advantage of using the RE criterion for G2 is that the RE representations form a primitive recursive set. Because of this, the question of whether a given expression of provability is RE is generally easier to answer than the question of whether it satisfies a given set of derivability conditions.

Using his new condition, Feferman proved the following version of G2 (1960: theorem 5.6).

If T is a system containing a minimal amount of arithmetic, $Ax_T(x)$ is an RE formula representing the axioms of T in T , and Con_T is built up from $Ax_T(x)$ in a way that corresponds to any of the usual ways in which the definition of consistency is related to the definition of axiomhood for T , then $\not\vdash_T \text{Con}_T$.

He also showed (theorem 5.9) that the restriction to RE cases of $Ax_T(x)$ is important since there are theories S and non-RE representations of their axioms $Ax_S^*(x)$ such that consistency formulas Con^*S - built up from $Ax_S^*(x)$ in exactly the same way that Con_T is built up from $Ax_T(x)$ - are provable in S . See Feferman (1982, 1989) for further development of his search for a general, readily applicable characterization of the representations of metamathematical notions with respect to which G2 holds.

Derivability conditions approaches have generally been enhanced by the recent development of provability logics (see **Provability logic**). The choice between the two types of approaches is complicated by the fact that Jeroslow's criterion for G2 appears to apply to theories to which Feferman's does not (namely, various cut-free systems of analysis - see Kreisel and Takeuti 1974). Thus, though Feferman's G2 criterion may generally be easier to apply than those of the derivability conditions approaches, it does not appear to be as broad in scope (see Feferman's remark 1 in Gödel (1990: 282-7) for a related discussion).

6 Philosophical applications

There have been many attempts to apply Gödel's incompleteness theorems to philosophical concerns. For the most part, these have to do with issues in the philosophy of mathematics and the philosophy of mind, though there have also been attempts in other areas of philosophy, and also in areas of intellectual life not lying wholly within philosophy (see Thomas 1995). Here we consider some of the better-known and/or better-developed applications.

In the philosophy of mathematics the chief application is to Hilbert's programme which is widely believed to have been refuted by G2. This is so because G2 is seen as prohibiting the finitary consistency proof that Hilbert demanded for classical mathematics (see, for example, Gödel 1958; Kitcher 1976; Kreisel 1958; Prawitz 1981; Resnik 1974; Smorynski 1977, 1985; Simpson 1988; for a dissenting view see Detlefsen 1986, 1990). G1 has been presented as having a similar effect through its exhibition of a recognizedly true real sentence of classical mathematics (namely, G) that is not finitarily provable (see Smorynski 1977, 1985; **Hilbert's programme and formalism §4**).

Gödel's theorems have also been used to argue against logicism. One example of this is in Hellman (1981), which maintains that G2 implies that no finitely axiomatizable logicist system exists. For non-finitely axiomatizable systems, the claim is weaker: such systems may exist, but G2 prohibits our being able to know of any particular system that it is one of them. A major attraction of Hellman's argument is that, unlike other arguments against logicism, it does not require that one know where to draw the line between logic and non-logic.

Gödel's theorems have also figured in philosophical discussions concerning the nature of proof. Among the more interesting of these are Myhill (1960) and Reinhardt (1985, 1986). Myhill argues that G1 and G2 establish that for any correct system containing arithmetic there are 'correct' inferences - inferences to whose acceptability the user of the system is rationally committed - that cannot be captured in it. There is thus, he says, an absolute epistemic notion of provability (neither syntactic, nor semantic nor psychological in character), according to which Gödel's undecidable sentences are provable. Hence, this notion of absolute provability is not formalizable. Reinhardt (1985, 1986) refines aspects of this argument and also argues that if 'humanly provable' means 'formally provable' there must be absolutely undecidable sentences of arithmetic. He urges the view that G2 is an epistemic phenomenon - one that arises from the epistemic properties of that type of belief that mathematical proof is supposed to sponsor. These properties, it turns out, lead to something tantamount to the derivability conditions.

In a related vein, Dummett (1963) has used G1 to argue that mathematical proof (including that which lies within a

well-circumscribed subject area) is an ‘indefinitely extensible concept’ and cannot therefore be identified with derivation in any formal system. Since, he believes, the meanings of mathematical statements are to be given in terms of a concept of proof, this means that the notion of proof that determines the meanings of mathematical statements must be an essentially vague and unformalizable one. In this he sounds a theme similar in certain respects to one associated with the intuitionist, Brouwer.

Chaitin (1974, 1975, 1982) offers an analysis of Gödel’s theorems that links them to limits on the extent to which information in a given mathematical field can be ‘compressed’ into a formal system. They thus, according to Chaitin, point up the need for a continuing search for new axioms (see [Computability and information](#)). This view of the relationship between Gödel’s theorems and algorithmic information theory is challenged in van Lambalgen (1989).

In the philosophy of mind and metaphysics, the main application of the theorems is to the question of mechanism (‘Do human-like minds have only such capabilities as are simulable by computational devices?’ - see [Mind, computational theories of §5](#)) and, more generally, to the question of materialism (‘Are all objects, events and/or forces in the world reducible to physical matter and its physical properties?’ - see [Materialism; Materialism in the philosophy of mind](#)). Gödel himself (1951) presented such an application, arguing that either the mind of the human mathematician cannot be codified by any formal system or that there exist absolutely unsolvable arithmetical problems of an elementary sort. On the former alternative, supposing Church’s thesis, mechanism clearly fails. On the latter, Gödel argued, materialism fails. For if there are mathematical problems that are absolutely unsolvable, then mathematics is not our own creation; and if this is so, its objects must exist independently of us, in which case a materialist view of reality fails. See Wang (1993) for a useful discussion of Gödel’s argument.

The most well-known anti-mechanist argument is that of Lucas (1961), modified by Benacerraf (1967) and repeated in Penrose (1989, 1994). In outline, the argument is as follows. (1) If human minds were mechanizable (or, in some versions, if they could be known to be identical to some particular machine), then, by G2, they could not know that their beliefs are consistent. But (2) human minds can know that their beliefs are consistent; therefore, (3) they are not mechanizable (or, as in Benacerraf’s variant, they cannot be known to be identical to any particular machine).

Lucas’ argument has been widely criticized (see, for example, Boolos 1968; Chihara 1972; Dennett 1978; Putnam 1960 (before the fact); Smart 1961; Wang 1974; Webb 1968, 1980). Its chief difficulty is perhaps one that is partially concealed by the unclarity of premise (2). It is true that humans are capable of examining various of their sets of beliefs and arriving at credible judgments regarding their consistency. It does not follow from this, however, that among the sets of beliefs they can judge to be consistent are some that contain that very consistency judgment itself. Indeed, if fixing a set of beliefs is a typical precondition for its evaluation, it would seem that consistency judgments do not typically belong to the sets to which they are applied; for at the time the beliefs are fixed such judgments would not yet have been made and so would not exist as elements of the set being evaluated. For consistency evaluations not requiring such prior fixing of beliefs, on the other hand, it is not clear that G2 applies. Either way, Lucas-type arguments face serious difficulties (see Detlefsen 1995).

See also: [Church’s theorem and the decision problem](#); [Church’s thesis](#); [Computability theory](#); [Logical and mathematical terms, glossary of](#); [Ordinal logics](#); [Proof theory](#)

MICHAEL DETLEFSEN

References and further reading

- Benacerraf, P.** (1967) ‘God, the Devil, and Gödel’, *The Monist* 51: 9-32. (A revised and chastened version of Lucas’ anti-mechanist argument from the incompleteness theorems.)
- Bernays, P.** (1926) ‘Axiomatische Untersuchung des Aussagen-Kalküls der *Principia Mathematica*’ (Axiomatic Investigation of the Propositional Calculus of *Principia Mathematica*), *Mathematische Zeitschrift* 25: 305-20. (Condensed version of the author’s 1918 *Habilitationsschrift* in which the first proof of the completeness of the propositional calculus of Whitehead and Russell was given.)
- Boolos, G.** (1968) ‘Review of “Minds, Machines and Gödel”, by J.R. Lucas, and “God, the Devil, and Gödel”, by P. Benacerraf’, *Journal of Symbolic Logic* 33: 613-15. (Critical review of Lucas (1961) and Benacerraf (1967).)

- Boolos, G. and Jeffrey, R.** (1974) *Computability and Logic*, Cambridge: Cambridge University Press, 3rd edn, 1989.(Thorough and well-written textbook of mathematical logic that provides much of the technical material required for the proofs given here.)
- Chaitin, G.J.** (1974) 'Information-Theoretic Limitations of Formal Systems', *Journal of the Association for Computing Machinery* 21: 403-24.(Introduction to Chaitin's analysis of Gödel's incompleteness theorems; argues that they show a limit on the information that can be 'compressed' into a formal axiomatic system, and infers that the search for new axioms must be an enduring part of mathematical research.)
- Chaitin, G.J.** (1975) 'Randomness and Mathematical Proof', *Scientific American* 232: 47-52.(Less technical version of various of the ideas in Chaitin (1974).)
- Chaitin, G.J.** (1982) 'Gödel's Theorem and Information', *International Journal of Theoretical Physics* 21: 941-54.(Attempt to explain Gödelian incompleteness as a phenomenon in algorithmic information theory.)
- Chihara, C.** (1972) 'On Alleged Refutations of Mechanism Using Gödel's Incompleteness Results', *Journal of Philosophy* 69: 507-26.(Critical discussion of attempted anti-mechanist applications of Gödel's incompleteness theorems.)
- Dedekind, R.** (1888) *Was sind und was sollen die Zahlen?*, Braunschweig: Vieweg; trans. W.W. Beman (1901), 'The Nature and Meaning of Numbers', in *Essays on the Theory of Numbers*, New York: Dover, 1963. (Includes Dedekind's proof of the categoricity of the second-order Peano postulates, chapter 10, theorem 132.)
- Dennett, D.C.** (1978) 'The Abilities of Men and Machines', in *Brainstorms: Philosophical Essays on Mind and Psychology*, Cambridge, MA: MIT Press, 1981.(Criticism of Lucas (1961) and similar arguments.)
- Detlefsen, M.** (1986) *Hilbert's Program: An Essay on Mathematical Instrumentalism*, Boston, MA, and Dordrecht: Reidel.(Detailed critical examination of the arguments from Gödel's incompleteness theorems against Hilbert's programme; focuses primarily on the arguments from G2 and the question of what is required of formulas that are to express consistency.)
- Detlefsen, M.** (1990) 'On an Alleged Refutation of Hilbert's Program using Gödel's First Incompleteness Theorem', *Journal of Philosophical Logic* 19: 343-77.(Critical examination of the arguments against Hilbert's programme from G1.)
- Detlefsen, M.** (1995) 'Wright on the Non-Mechanizability of Intuitionist Reasoning', *Philosophia Mathematica*, series 3, 3: 103-19.(Critical examination of an argument by Crispin Wright that Gödel's theorems imply the non-mechanizability of intuitionist mathematics. Includes brief statement of what the author regards as the central mistake of all Lucas-type arguments.)
- Dummett, M.A.E.** (1963) 'The Philosophical Significance of Gödel's Theorem', in *Truth and Other Enigmas*, Cambridge, MA: Harvard University Press, 1978.(Includes Dummett's argument that Gödel's incompleteness theorems show that the notion of mathematical proof is essentially unformalizable.)
- Enderton, H.B.** (1972) *A Mathematical Introduction to Logic*, New York: Academic Press, 1972. (Well-conceived and -executed basic textbook of mathematical logic; supplies definitions and proofs of the basic notions and results needed to prove G1. Treatment of G2 is sketchy.)
- Feferman, S.** (1957) 'Formal Consistency Proofs and Interpretability of Theories', in *Summaries of Talks Presented at the Summer Institute of Symbolic Logic in 1957 at Cornell University*, vol. 1, 71-7. Mimeographed.(A succinct and readable summary of the author's Berkeley doctoral dissertation.)
- Feferman, S.** (1960) 'Arithmetization of Metamathematics in a General Setting', *Fundamenta Mathematicae* 49: 3-92.(Thorough treatment of the differences between G1 and G2 as regards the conditions they require of arithmetical expressions of metamathematical notions. Develops a novel and interesting alternative to the traditional derivability conditions approach to the proof of G2.)
- Feferman, S.** (1982) 'Inductively Presented Systems and the Formalization of Metamathematics', in D. van Dalen, D. Lascar and T. Smiley (eds) *Logic Colloquium '80*, Studies in Logic and the Foundations of Mathematics, 108, Amsterdam: North Holland, 1982.(Outline of investigation designed to extend and refine the study begun in the author's dissertation and 1960 paper. The aim is a fully explicit and precise account of canonical representations of the various metamathematical notions pertaining to formal axiomatic systems.)
- Feferman, S.** (1989) 'Finitary Inductively Presented Logics', in R. Ferro, Bonotto, S. Valentini and A. Zanardo (eds) *Logic Colloquium '88*, Studies in Logic and the Foundations of Mathematics, 127, Amsterdam: North Holland, 1989.(Continuation of the author's long-standing study of what is required for the canonical representation of formal systems.)
- Gauss, K.** (1817, 1829) letter to Olbers (1817), in *Briefwechsel mit H.W.M. Olbers*, Hildesheim: Olms, 1976; letter to Bessel (1829), in *Werke*, Leipzig: Teubner, 1863-1903, vol. 8, 200.(Statement of Gauss' belief that though a human knower might be able to attain complete knowledge of the laws of arithmetic a priori, the same would not be true of the laws of geometry.)

- Gödel, K.** (1929) 'Über die Vollständigkeit des Logikkalküls', doctoral dissertation, University of Vienna; trans. 'On the Completeness of the Logical Calculus', in Gödel (1986), 44-101; slightly modified version, 'Die Vollständigkeit der Axiome des logischen Funktionenkalküls', *Monatshefte für Mathematik und Physik* 37: 349-60, 1930; trans. 'The Completeness of the Axioms of the Functional Calculus of Logic', in van Heijenoort (1967), 582-91; and in Gödel (1986), 102-23. (Gödel's doctoral dissertation in which he first proved the completeness of the quantificational portion of the logic of Whitehead and Russell.)
- Gödel, K.** (1930) 'Einige metamathematische Resultate über Entscheidungsdefintheit und Widerspruchsfreiheit', repr. and trans. 'Some Metamathematical Results on Decidability and Consistency', in Gödel (1986); trans. in van Heijenoort (1967). (Brief announcement of Gödel's incompleteness theorems.)
- Gödel, K.** (1931) 'Über formal unentscheidbare Sätze der *Principia Mathematica* und verwandter Systeme I', *Monatshefte für Mathematik und Physik* 38: 173-98; trans. 'On Formally Undecidable Propositions of *Principia Mathematica* and Related Systems', in van Heijenoort (1967), 592-617; and in Gödel (1986), 144-95. (Paper in which were originally published Gödel's proofs of his incompleteness theorems.)
- Gödel, K.** (1951) 'Some Basic Theorems on the Foundations of Mathematics and Their Implications', in Gödel (1995). (Edited text of Gödel's 1951 Josiah Gibbs Lecture delivered at Brown University at a meeting of the American Mathematical Society. Gödel here formulated his argument concerning the anti-materialist implications of his incompleteness theorems; also includes an argument for a Platonist conception and an argument against a conventionalist conception of mathematics. Useful introductory essay by George Boolos.)
- Gödel, K.** (1958) 'Über eine bisher noch nicht benützte Erweiterung des finiten Standpunktes', repr. and trans. 'On a Hitherto Unexploited Extension of the Finitary Standpoint', in Gödel (1990). (Gödel's attempt to find a legitimate epistemological extension of Hilbert's finitary standpoint that is capable of providing the kinds of soundness proofs for classical mathematics that Hilbert called for. Includes Gödel's statement that, unlike in his 1931 paper, he now accepted the idea that G2 essentially refutes Hilbert's programme.)
- Gödel, K.** (1986-95) *Collected Works*, ed. S. Feferman et al., vol. 1, *Publications 1929-1936*, 1986; vol. 2, *Publications 1938-1974*, 1990; vol. 3, *Unpublished Essays and Lectures*, 1995, New York and Oxford: Oxford University Press. German originals and English translations. (Accurate yet readable translations, useful introductory essays and bibliography.)
- Hasenjäger, G.** (1953) 'Eine Bemerkung zu Henkin's Beweis für die Vollständigkeit des Prädikatenkalküls der ersten Stufe' (A Remark on Henkin's Proof of the Completeness of the First-Order Predicate Calculus), *Journal of Symbolic Logic* 18: 42-8. (A simplification of Henkin's completeness proof.)
- Heijenoort, J. van** (ed.) (1967) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press. (Basic papers in mathematical logic and the foundations of mathematics, with useful forewords and bibliography.)
- Hellman, G.** (1981) 'How to Gödel a Frege-Russell: Gödel's Incompleteness Theorems and Logicism', *Noûs* 25: 451-68. (Argues that G2 counts against logicism.)
- Henkin, L.** (1949) 'The Completeness of the First-Order Functional Calculus', *Journal of Symbolic Logic* 14: 159-66. (First presentation of the currently prevailing method for completeness proofs; introduced devices making the proof less dependent upon the details of the particular calculus involved and, hence, more general in scope.)
- Henkin, L.** (1952) 'A Problem concerning Provability', *Journal of Symbolic Logic* 17: 160. (Posed the question 'Is a sentence expressing the statement that it is itself provable in *T*, a first-order formal system containing enough arithmetic to enable it to represent all recursive relations of natural numbers, provable in *T* or not?' Solved in the affirmative by Löb (1955).)
- Hilbert, D.** (1900) 'Über den Zahlbegriff' (On the Number-Concept), *Jahresbericht der Deutschen Mathematiker-Vereinigung* 8; revised as appendix 6 of *Grundlagen der Geometrie*, Leipzig and Berlin: Teubner, 7th edn, 1930. (Early statement by Hilbert of the chief aims of foundational studies. Explicitly identified are a completeness problem, and the consistency problem, which is described as necessary for the demonstration of the existence of a concept.)
- Hilbert, D. and Ackermann, W.** (1928) *Grundzüge der theoretischen Logik*, Berlin: Springer, 2nd edn, 1938; 2nd edn trans. L.M. Hammond, G.G. Leckie and F. Steinhardt, *Principles of Mathematical Logic*, ed. R.E. Luce, New York: Chelsea, 1950. (Text in which the standard problems of metamathematics - consistency, completeness, independence and decidability - received their first precise formulations; readable and illuminating source for students today. These formulations guided Gödel and Church in their subsequent research.)
- Hilbert, D. and Bernays, P.** (1939) *Grundlagen der Mathematik (Foundations of Mathematics)*, vol. 2, Berlin: Springer; 2nd edn, 1970. (Classic work of mathematical logic; gives the first full proof of G2.)

- Hintikka, J.** (1955) 'Form and Content in Quantification Theory', *Acta Philosophica Fennica* 8: 7-55.
(Description of a tableau method that supports a simple proof of completeness.)
- Huntington, E.V.** (1905) 'A Set of Postulates for Real Algebra, Comprising Postulates for a One-Dimensional Continuum and for the Theory of Groups', *Transactions of the American Mathematical Society* 6: 17-41.
(Includes early explicit mention of completeness as a property of an axiom system (see page 17).)
- Jeroslow, R.** (1973) 'Redundancies in the Hilbert-Bernays Derivability Conditions for Gödel's Second Incompleteness Theorem', *Journal of Symbolic Logic* 38: 359-67.(Shows how to eliminate DC2 as a condition on provability expressions whose associated consistency expressions support a version of G2.)
- Kitcher, P.** (1976) 'Hilbert's Epistemology', *Philosophy of Science* 43: 99-115.(Useful discussion of the epistemic character of Hilbert's so-called 'finitary' judgments and of the bearing of Gödel's theorems on Hilbert's programme.)
- Kreisel, G.** (1958) 'Hilbert's Programme', *Dialectica* 12: 346-72; revised version in P. Benacerraf and H. Putnam (eds) *Philosophy of Mathematics: Selected Readings*, Cambridge: Cambridge University Press, 2nd edn, 1983.
(Essay on the plausibility of Hilbert's programme in light of Gödel's theorems and other developments; discusses the differences between G1 and G2 and between what Kreisel calls the broader and narrower versions of Hilbert's programme; claims that G1 counts against the broader version.)
- Kreisel, G.** (1965) 'Mathematical Logic', in T.L. Saaty (ed.) *Lectures in Modern Mathematics*, vol. 3, New York: Wiley.(Survey of modern mathematical logic. Considers the conditions that must be satisfied by formulas expressing metamathematical notions, and introduces the notion of a 'canonical' representation.)
- Kreisel, G. and Takeuti, G.** (1974) 'Formally Self-Referential Propositions for Cut-Free Classical Analysis and Related Systems', *Dissertationes Mathematicae* 118: 4-50.(Useful study of the conditions on formulas expressing provability that are required for G2 and such related results as Löb's theorem. Shows how the streamlining of the derivability conditions discovered by Jeroslow allows a proof of G2 for certain cut-free systems of analysis.)
- Kronecker, L.** (1887) 'Über den Zahlbegriff' (On the Concept of Number), *Journal für die reine und angewandte Mathematik* 101: 337-55; repr. in *Werke*, ed. K. Hensel, Leipzig: Teubner, 1932, vol. 3, part 1, 249-74.
(Kronecker's only philosophical essay, in which he expresses the view, also expressed earlier by Gauss, that although humans might be able to attain a complete knowledge of the laws of arithmetic a priori, the same cannot be expected of the laws of geometry.)
- Lambalgem, M. van** (1989) 'Algorithmic Information Theory', *Journal of Symbolic Logic* 54: 1,389-400.(Argues that Chaitin's analysis of Gödel's theorem in terms of algorithmic information theory does not in fact explain it.)
- Löb, M.H.** (1955) 'Solution of a Problem of Leon Henkin', *Journal of Symbolic Logic* 20: 115-18. (Positive solution of the problem posed by Henkin (1952). Streamlined the original conditions on provability expressions used in the first full proof of G2 in Hilbert and Bernays (1939).)
- Löwenheim, L.** (1915) 'Über Möglichkeiten im Relativkalkül', *Mathematische Annalen* 76: 447-70; trans. S. Bauer-Mengelberg, 'On Possibilities in the Calculus of Relatives', in van Heijenoort (1967), 232-51.(Includes a proof of Löwenheim's theorem that a formula true in all finite and denumerably infinite domains is true in all domains. Makes use of a normal form for first-order formulas that Gödel also made use of in his completeness proof.)
- Lucas, J.R.** (1961) 'Minds, Machines, and Gödel', *Philosophy* 36: 112-27; revised version appears in *Freedom of the Will*, Oxford: Clarendon Press, 1970.(First statement of a well-known and widely discussed argument that Gödel's incompleteness theorems refute mechanism.)
- Lucas, J.R.** (1970) *Freedom of the Will*, Oxford: Clarendon Press.(Includes responses to some of the criticisms of the earlier version of the argument (1961) and gives a select bibliography of the anti-mechanist argument from Gödel's theorems.)
- Maltsev, A.I.** (1936) 'Untersuchungen aus dem Gebiete der mathematischen Logik', *Matematicheskii Sbornik* 1: 323-6; trans. 'Investigations in the Area of Mathematical Logic', in B. Wells (ed.) *The Metamathematics of Algebraic Systems: Collected Papers, 1936-1967*, Amsterdam: North Holland, 1971.(Extends to uncountable sets of formulas the proof of compactness given in Gödel (1930) for countable sets of formulas.)
- Myhill, J.** (1960) 'Some Remarks on the Notion of Proof', *Journal of Philosophy* 62: 461-70.(Argues that Gödel's incompleteness theorems show that for any correct formal system containing arithmetic there are inferences to whose acceptability the user of the system is rationally committed that cannot be captured in it.)
- Paris, J. and Harrington, L.** (1977) 'A Mathematical Incompleteness in Peano Arithmetic', in J. Barwise (ed.) *Handbook of Mathematical Logic*, Amsterdam: North Holland.(First proof of a reasonably natural and true statement of pure combinatorics that is not provable in first-order Peano arithmetic. Followed a few years later

- by an even more striking discovery of Harvey Friedman's (see Simpson 1985).)
- Penrose, R.** (1989) *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics*, New York: Oxford University Press. (Argues against mechanism in a manner essentially that of Lucas (1961).)
- Penrose, R.** (1994) *Shadows of the Mind: A Search for the Missing Science of Consciousness*, New York: Oxford University Press. (Includes replies to critics of the anti-mechanist argument of Penrose (1989). Essential elements of the argument remain unchanged.)
- Post, E.** (1921) 'Introduction to a General Theory of Elementary Propositions', *American Journal of Mathematics* 43: 163-85; repr. in van Heijenoort (1967), 264-83. (First published proof of the completeness of the propositional calculus of Whitehead and Russell; also proved its Post-completeness.)
- Prawitz, D.** (1981) 'Philosophical Aspects of Proof Theory', in G. Fløistad (ed.) *Contemporary Philosophy: A New Survey*, vol. 1, The Hague: Martinus Nijhoff, 1981. (Useful presentation of the case against Hilbert's programme via Gödel's theorems, especially G1.)
- Putnam, H.** (1960) 'Minds and Machines', in S. Hook (ed.) *Dimensions of Mind: A Symposium*, New York: Macmillan. (Predates Lucas (1961), but none the less has the makings of an effective argument against it.)
- Reinhardt, W.** (1985) 'Absolute Versions of Incompleteness Theorems', *Noûs* 19: 317-46. (Argues that epistemic properties of an intuitive notion of provability give rise to G1 and G2 phenomena.)
- Reinhardt, W.** (1986) 'Epistemic Theories and the Interpretation of Gödel's Incompleteness Theorems', *Journal of Philosophical Logic* 15: 427-74. (Extends and refines the discussion of Reinhardt (1985).)
- Resnik, M.** (1974) 'The Philosophical Significance of Consistency Proofs', *Journal of Philosophical Logic* 3: 133-47. (Discussion of certain elements of Feferman (1960) and the question of the difference between normal and 'pathological' expressions of metamathematical notions.)
- Rosser, J.B.** (1936) 'Extensions of Some Theorems of Gödel and Church', *Journal of Symbolic Logic* 1: 87-91. (Proof of a version of G1 that does not rely upon a theory's being ω -consistent.)
- Simpson, S.** (1985) 'Nonprovability of certain combinatorial properties of finite trees', in L. Harrington *et al.* (eds) *Harvey Friedman's Research in the Foundations of Mathematics*, Amsterdam: North Holland, 1985. (First published presentation of Friedman's proof in the early 1980s that a certain proposition of finite combinatorics (a form of Kruskal's theorem on finite trees) is independent not only of first-order Peano arithmetic but also of two subsystems of second-order Peano arithmetic that go well beyond it.)
- Simpson, S.** (1988) 'Partial Realization of Hilbert's Program', *Journal of Symbolic Logic* 53: 349-63. (Useful survey and basic exposition of the work of the so-called 'reverse mathematics' programme begun by Harvey Friedman. Argues that though Hilbert's programme in its original form is refuted by G2, a significant portion of that programme can still be carried out.)
- Skolem, T.** (1923) 'Einige Bemerkungen zur axiomatischen Begründung der Mengenlehre', in *Matematikerkongressen i Helsingfors den 4-7 Juli 1922, Den femte skandinaviska matematikerkongressen, Redogörelse*, Helsinki: Akademiska Bokhandeln, 217-32; trans. S. Bauer-Mengelberg, 'Some Remarks on Axiomatized Set Theory', in van Heijenoort (1967), 291-301. (Anticipated the chief result of Gödel (1930). Gödel was apparently unaware of it at the time he wrote his paper.)
- Smart, J.J.C.** (1961) 'Gödel's Theorem, Church's Theorem and Mechanism', *Synthese* 13: 105-10. (Critical discussion of Lucas-type arguments.)
- Smoryński, C.** (1977) 'The Incompleteness Theorems', in J. Barwise (ed.) *Handbook of Mathematical Logic*, Amsterdam, North Holland. (Exposition of Gödel's incompleteness theorems, their proofs and some of their applications.)
- Smoryński, C.** (1985) *Self-Reference and Modal Logic*, New York: Springer. (Good exposition of provability logic. Chapter 0 presents argument against Hilbert's programme via G1.)
- Smullyan, R.** (1961) *Theory of Formal Systems*, Princeton, NJ: Princeton University Press. (Important study of the nature of formal systems and what is required for the adequate representation of metamathematical notions via arithmetization.)
- Thomas, D.** (1995) 'Gödel's Theorem and Postmodern Theory', *Publications of the Modern Language Association* 110: 248-61. (Discussion of the significance of Gödel's incompleteness theorems for literary criticism.)
- Wang, H.** (1974) *From Mathematics to Philosophy*, New York: Humanities Press. (Includes worthwhile discussion of a variety of aspects of Gödel's theorems and of his philosophical ideas; also a critical discussion of Lucas-type arguments against mechanism.)
- Wang, H.** (1993) 'Can Machines Think?', *Philosophia Mathematica*, series 3, 1: 97-138. (Discussion of Gödel's ideas on mechanism. Reconstructs the many conversations between the author and Gödel on the subject.)
- Webb, J.** (1968) 'Metamathematics and the Philosophy of Mind', *Philosophy of Science* 35: 156-78. (Interesting

discussion of some of the difficulties involved in trying to link Gödel's incompleteness theorems to anti-mechanist views of mind.)

Webb, J. (1980) *Mechanism, Mentalism, and Metamathematics: An Essay on Finitism*, Boston, MA: Reidel.

(Thorough study of various themes in philosophy with which Gödel's incompleteness theorems have sometimes been associated. Argues that, contrary to the opinion of some, these theorems actually confirm a type of formalism in the philosophy of mathematics.)

Whitehead, A.N. and Russell, B.A.W. (1910) *Principia Mathematica*, vol. 1, Cambridge: Cambridge University Press; 2nd edn, 1925; repr. 1994.(The classical symbolic formalization of logic and mathematics; basis of much of the greatest work in mathematical logic and the foundations of mathematics in the twentieth century.)

Herbrand's theorem

According to Herbrand's theorem, each formula F of quantification theory can be associated with a sequence $F_1, F_2, F_3, \dots$ of quantifier-free formulas such that F is provable just in case F_n is truth-functionally valid for some n . This theorem was the centrepiece of Herbrand's dissertation, written in 1929 as a contribution to Hilbert's programme. It provides a finitistically meaningful interpretation of quantification over an infinite domain. Furthermore, it can be applied to yield various consistency and decidability results for formal systems. Herbrand was the first to exploit it in this way, and his work has influenced subsequent research in these areas. While Herbrand's approach to proof theory has perhaps been overshadowed by the tradition which derives from Gentzen, recent work on automated reasoning continues to draw on his ideas.

Herbrand's theorem says that each formula F of quantification theory can be associated with a sequence $F_1, F_2, F_3, \dots$ of quantifier-free formulas such that F is provable just in case F_n is truth-functionally valid for some n . If F is prenex - that is, consists of a quantifier-free matrix G preceded by a string of quantifiers - then each F_i will be a disjunction of instances of G obtained by substituting terms for free variables. To simplify the exposition we consider only prenex formulas F . With each such F we can associate an existential formula F^H as follows: (1) substitute new constants for the free variables of F ; (2) remove each universal quantifier $\forall x$ and replace the resulting free occurrences of x by a term $f(y_1, \dots, y_m)$, where f is a new function symbol and $y_1, \dots, y_m$ are all the variables bound by existential quantifiers to the left of $\forall x$ in F . For example, applied to $\forall x \exists y \forall z R(x, y, z)$ (with R quantifier-free) this procedure yields $\exists y R(c, y, f(y))$.

Although Herbrand would not have described their relationship in these terms, F^H can be said to express the meaning of F in the sense that F is true under an interpretation $\mathbb{I}$ just in case F^H is true under any expansion of $\mathbb{I}$ to the new constants and function symbols in F^H . To see this, consider our example. Let α, β range over the domain of $\mathbb{I}$, ϕ over functions on this domain and let ' x_α ' indicate that the value α is assigned to the variable x . Then the following are equivalent:

$\forall x \exists y \forall z R(x, y, z)$ is true under $\mathbb{I}$

There is no α such that for all $\beta \exists z \neg R(x_\alpha, y_\beta, z)$.

There is no α and no function ϕ such that for all $\beta \neg R(x_\alpha, y_\beta, z_{\phi(\beta)})$.

For every α and every function ϕ there is some β such that $R(x_\alpha, y_\beta, z_{\phi(\beta)})$.

No matter which individual α is chosen to interpret c and which function ϕ to interpret f , $\exists y R(c, y, f(y))$ is true under $\mathbb{I}$.

This means, in particular, that F is true under every interpretation iff F^H is; the latter is called the validity functional form of F .

We can use the functional terms appearing in F^H to generate a series of finite domains $\mathbf{D}_i$. These will consist of constants which can themselves be substituted into the quantifier-free matrix G^H of F^H . For convenience, let $\mathbf{D}_1$ consist of a single constant d , $\mathbf{D}_2$ of a distinct constant (its 'denotation') for each term obtained from one occurring in G^H by substituting d for its free variables; in particular, constants occurring in G^H are assigned denotations in $\mathbf{D}_2$. In general, $\mathbf{D}_{n+1}$ consists of distinct denotations for each term obtained from one occurring in G^H by substituting members of $\mathbf{D}_n$ for its free variables. F_n , the Herbrand expansion of F of order n , is a disjunction whose disjuncts represent all possible ways of substituting members of $\mathbf{D}_n$ for the variables of G^H and replacing functional terms, where possible, by their values in $\mathbf{D}_n$. For example, if we allow each closed term to denote itself, then the first three domains generated by our earlier formula will be $\{d\}$, $\{d, c, f(d)\}$ and $\{d, c, f(d), f(c), f(f(d))\}$. The Herbrand expansions of orders 1, 2 and 3 are:

$$\begin{aligned} & R(c, d, f(d)) \\ & R(c, d, f(d)) \vee R(c, c, f(c)) \vee R(c, f(d), f(f(d))) \\ & R(c, d, f(d)) \vee R(c, c, f(c)) \vee R(c, f(d), f(f(d))) \\ & \quad \vee R(c, f(c), f(f(c))) \vee R(c, f(f(d)), f(f(f(d)))) \end{aligned}$$

Notice that F_n is not an interpretation of the quantifiers of F^H in $\mathbf{D}_n$ in the usual sense because some of the

terms occurring in F_n only acquire a denotation in $\mathbf{D}_{n+1}$.

Now, for provable F , Herbrand showed how to compute a number n such that F_n would be a tautology. He argued by induction on the proof of F . The only axioms in his system are the quantifier-free tautologies, so his method is to analyse each rule of inference and show how the order of a tautologous expansion of its conclusion can be calculated from the order(s) of such expansions of its premises. Conversely, he showed how to construct a proof of F from any tautologous F_n . Because the only rules required here are forms of generalization and simplification (in particular, *modus ponens* is not used), this yields a normal form for proofs in quantification theory analogous to Gentzen's cut-elimination theorem and lends itself to many of the same applications. It allows a bound to be placed on the complexity of formulas appearing in a normal proof depending on the complexity of its conclusion (and assumptions, if any).

Herbrand's proof contains ideas of great interest but, in contrast to Gentzen's, it is not easily adapted to non-classical systems. Furthermore, it is difficult to follow and turns out to contain a defect. Although this has been repaired by Dreben and his collaborators, the proof has not been reproduced and nowadays the theorem (for prenex formulas only) is often established as a corollary to a refinement of the cut-elimination theorem. (Since a cut-free proof of a prenex formula can be separated into a truth-functional and a quantificational part, by working back from the conclusion we can discover a truth-functionally valid formula from which it follows.)

While he accepted set-theoretic arguments in mathematics, Herbrand was committed to using finitary methods in metamathematical investigations. His work was influenced by Löwenheim (1915), but he rejected the naïve notion of satisfiability in an infinite domain the latter assumed. In fact, Herbrand claimed that his result was just a more rigorous version of Löwenheim's famous theorem to the effect that a sentence is satisfiable if and only if it is satisfiable in a denumerable domain. It would be more accurate, however, to describe it as a finitist variant thereof. To understand why this is so, notice that we can reformulate Herbrand's theorem using a construction dual to the above. We remove existential quantifiers from F , replacing the newly freed variables by functional terms to obtain a universal formula (the satisfiability functional form of F), which can then be expanded as a series of finite conjunctions F'_n over the $\mathbf{D}_n$. The theorem then states that F is irrefutable iff F'_n is truth-functionally satisfiable for every n . (This is the original theorem expressed as a criterion for the non-provability of $\neg F$.) Constrained by his finitist scruples, Herbrand explicated validity in terms of provability and, *a fortiori*, satisfiability in terms of irrefutability. He also suggested that F 's having truth-functionally satisfiable expansions F'_n for every n could serve as a constructive account of what it means for F to be satisfiable in a (denumerable) domain. Replacing the informal set-theoretic notions in the statement of Löwenheim's theorem by their finitist analogues transforms it into a statement of Herbrand's theorem - even though the theorems themselves remain distinct.

Herbrand's ideas have yielded numerous metamathematical applications. The normal form for proofs mentioned above supplies a tool for obtaining consistency proofs. Furthermore, the elimination of quantifiers in favour of functional terms suggests how to exhibit the constructive content of the theorems of a formal theory - as in the 'no counterexample' interpretation of arithmetic, for example (Kreisel 1951). Herbrand's theorem also allows us to prove a theory consistent by approximating a model for it, that is, by establishing that, for every (finite) conjunction F of the theory's axioms, F'_n is truth-functionally satisfiable for every n . Herbrand himself gave a consistency proof for a fragment of arithmetic along these lines, and his methods have been extended to yield proofs for the full system (see Dreben and Denton 1970, Scanlon 1972). These proofs, while no less complex than Gentzen's, yield directly the sort of constructive interpretations alluded to above.

Finally, Herbrand's theorem supplies an approach to the decision problem for classes of quantificational formulas, that is, to discovering an algorithm which, for any formula in the class, computes whether or not it is satisfiable. The idea here is to investigate the relationship between syntactic properties of quantificational formulas and the structure of their expansions, and to discover properties of the latter which ensure the existence of a decision procedure (see Dreben and Goldfarb 1979). His theorem and the methods introduced in its proof are relevant not only to theoretical decidability results, but also to the design of implementable automated reasoning procedures. These employ unification algorithms to improve the efficiency of testing quantifier-free expansions for satisfiability, and Herbrand is credited with being the first to devise such an algorithm (Snyder 1991).

See also: Hilbert's programme and formalism; Logical and Mathematical Terms, Glossary of; Proof theory

A.M. UNGAR

References and further reading

- Andrews, P.B.** (1986) *An Introduction to Mathematical Logic and Type Theory: To Truth Through Proof*, New York: Academic Press. (§35 contains an interesting discussion of Herbrand's theorem, focusing on its significance for decision problems and automated theorem proving, as well as a proof of one version of it.)
- Buss, S.R.** (1995) 'On Herbrand's Theorem', in D. Leivant (ed.) *Logic and Computational Complexity*, Berlin: Springer, 195-209. (An up-to-date exposition of the theorem from a proof-theoretic perspective which also sets it in the context of Herbrand's dissertation.)
- Dreben, B. and Denton, J.** (1970) 'Herbrand-Style Consistency Proofs', in J. Myhill, R.E. Vesley and A. Kino (eds) *Intuitionism and Proof Theory*, Amsterdam: North Holland, 419-33. (Building on ideas to be found in Herbrand, the authors sketch a proof of the consistency of arithmetic and discuss some of its corollaries.)
- Dreben, B. and Goldfarb, W.D.** (1979) *The Decision Problem: Solvable Classes of Quantificational Formulas*, Reading, MA: Addison-Wesley. (This monograph offers a unified treatment of solvable cases of the decision problem by way of an approach that can be traced back to Herbrand.)
- Goldfarb, W.D.** (1979) 'Logic in the Twenties: The Nature of the Quantifier', *Journal of Symbolic Logic* 44 (3): 351-68. (A very readable discussion of the background to Herbrand's work.)
- Heijenoort, J. van** (1985) 'Jacques Herbrand's Work in Logic and its Historical Context', in *Selected Essays*, Naples: Bibliopolis, 99-121. (A readable informal survey of Herbrand's work.)
- Herbrand, J.** (1968) *Écrits logiques*, Paris: Presses Universitaires de France; trans. *Logical Writings*, ed. W.D. Goldfarb, Cambridge, MA: Harvard University Press, 1971. (All of Herbrand's logical writings, with extensive annotations, needed corrections and a helpful introduction by the editor.)
- Hilbert, D. and Bernays, P.** (1939) *Grundlagen der Mathematik (Foundations of Mathematics)*, vol. 2, Berlin: Springer, 2nd edn, 1970. (A standard reference; Herbrand's theorem is derived from the first ε -elimination theorem and its metamathematical applications detailed.)
- Kreisel, G.** (1951, 1952) 'On the Interpretation of Nonfinitist Proofs', *Journal of Symbolic Logic* 16 (4): 241-67, 17 (1): 43-58. (The methods of Hilbert and Bernays (1939) are utilized to impart a finitist sense to proofs in arithmetic and a subsystem of analysis.)
- Löwenheim, L.** (1915) 'Über Möglichkeiten im Relativkalkül', *Mathematische Annalen* 76: 447-70; trans. S. Bauer-Mengelberg, 'On Possibilities in the Calculus of Relatives', in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967. (A seminal paper which, because its results are stated and proved in terms of Schröder's calculus of relatives, is accessible only with difficulty to the modern reader.)
- Scanlon, T.M.** (1972) 'The Consistency of Number Theory via Herbrand's Theorem', *Journal of Symbolic Logic* 38 (1): 29-58. (This paper presents in some detail an improved and slightly more general version of the proof given by Dreben and Denton (1970).)
- Schwichtenberg, H.** (1977) 'Proof Theory: Some Applications of Cut-Elimination', in J. Barwise (ed.) *Handbook of Mathematical Logic*, Amsterdam: North Holland, 867-95. (Herbrand's theorem is derived from the cut-elimination theorem and some of its proof theoretic applications are sketched.)
- Shoenfield, J.** (1967) *Mathematical Logic*, Reading, MA: Addison-Wesley. (Chapter 4 includes an accessible and enlightening proof of Herbrand's theorem for the predicate calculus with equality.)
- Snyder, W.** (1991) *A Proof Theory for General Unification*, Boston, MA: Birkhäuser. (A specialized monograph which describes an abstract framework within which to understand different kinds of unification.)

Hilbert's programme and formalism

In the first, geometric stage of Hilbert's formalism, his view was that a system of axioms does not express truths particular to a given subject matter but rather expresses a network of logical relations that can (and, ideally, will) be common to other subject matters.

The formalism of Hilbert's arithmetical period extended this view by emptying even the logical terms of contentual meaning. They were treated purely as ideal elements whose purpose was to secure a simple and perspicuous logic for arithmetical reasoning - specifically, a logic preserving the classical patterns of logical inference. Hilbert believed, however, that the use of ideal elements should not lead to inconsistencies. He thus undertook to prove the consistency of ideal arithmetic with its contentual or finitary counterpart and to do so by purely finitary means.

In this, 'Hilbert's programme', Hilbert and his followers were unsuccessful. Work published by Kurt Gödel in 1931 suggested that such failure was perhaps inevitable. In his second incompleteness theorem, Gödel showed that for any consistent formal axiomatic system T strong enough to formalize what was traditionally regarded as finitary reasoning, it is possible to define a sentence that expresses the consistency of T , and is not provable in T . From this it has generally been concluded that the consistency of even the ideal arithmetic of the natural numbers is not finitarily provable and that Hilbert's programme must therefore fail.

Despite problematic elements in this reasoning, post-Gödelian work on Hilbert's programme has generally accepted it and attempted to minimize its effects by proposing various modifications of Hilbert's programme. These have generally taken one of three forms: attempts to extend Hilbert's finitism to stronger constructivist bases capable of proving more than is provable by strictly finitary means; attempts to show that for a significant family of ideal systems there are ways of 'reducing' their consistency problems to those of theories possessing more elementary (if not altogether finitary) justifications; and attempts by the so-called 'reverse mathematics' school to show that the traditionally identified ideal theories do not need to be as strong as they are in order to serve their mathematical purposes. They can therefore be reduced to weaker theories whose consistency problems are more amenable to constructivist (indeed, finitist) treatment.

1 The formalism of the geometric period

Hilbert's conception of mathematics during the period of his work on the foundations of geometry was 'formalist' in the sense that it viewed an axiomatic system as providing not a family of truths specific to a particular subject matter but, rather, a 'theory-form' or hypothetico-deductive framework which potentially and, indeed, preferably served as the form for a variety of different particular sciences.

This doctrine has two distinct moments that it is important to distinguish. One is a 'rigourist' moment which maintains that, for the sake of achieving rigour in mathematical reasoning, axiomatic systems must be formalized to the point of emptying their distinctively mathematical concepts of all intuitional or distinctly mathematical content so that they become mere 'positions' in a network of logical relations. The other moment is a 'schematist' one which says that even for the sake of achieving truth in one's axioms, the non-logical terms in an axiomatic theory do not have to be tied to any particular intuitive content. Conceived in this way, the mathematical terms of an axiomatic theory become 'variables' of a sort (that is, terms that range over a variety of different intuitive contents) rather than 'constants' (that is, terms that signify a unique particular content).

Historically, the rigourist moment had its origins in the seventeenth-century discovery that various proofs in Euclid's *Elements* relied upon assumptions not stated in his axioms, postulates and definitions. In the eighteenth century, Lambert (1766: 162) proposed a remedy for such failures of rigour. He reasoned that since they occur when one makes tacit use of geometric intuition while conducting an inference in a proof, they ought to be eradicable by forcing all inferences to be validated without appeal to geometric intuition. This, he maintained, could be accomplished by requiring the validation of inferences to proceed in abstraction from the intuitive meanings of the mathematical terms involved and to be warranted solely on the basis of their 'symbolic characteristics'. Axioms would thus be treated like 'algebraic equations' in which mathematical terms are manipulated according to the logical positions laid down for them in the axioms and not according to their intuitive meanings.

This, essentially, was the conception of rigour that Pasch (1882) and Hilbert (1899a) put into effect in their axiomatizations of geometry a century later.

Inference, however, is only one part of mathematical practice and rigour, therefore, only one of its goals. Traditionally, it has also been required that the theorems, and, hence, the axioms, be true. To satisfy this goal, however, would seem to require treating the mathematical terms occurring in the axioms from the point of view of their intuitive meanings or contents. Such, at any rate, is the traditional view. The schematist moment of Hilbert's formalism challenges this view by maintaining that, ideally, a system of axioms should have a number of different (important) realizations and that these different realizations should be obtainable from one another by relatively simple schemes of interchange of the meanings of the non-logical terms in the axioms. What is most important, then, is for a system of axioms to have not one particular realization but a variety of different ones.

This interesting view is sometimes distorted and misrepresented by saying that Hilbert defined truth as consistency or that he advocated mere consistency rather than truth (as ordinarily defined) as the goal of mathematical theorizing. Hilbert sometimes said things that encourage such readings (1899b: 39-40, 1900: 448, 1905: 125, 1922: 157). However, more careful consideration of the ideas that influenced his views together with a balanced reading of the texts themselves (1922: 162-5, 1923: 179-80; Hilbert and Bernays 1934: §1) indicates that this was not his position.

This type of formalist view was suggested by a striking phenomenon of nineteenth-century geometry - the so-called 'dualities' that emerged in projective geometry. Generally speaking, a duality is a pair of theorems one of which can be obtained from the other by a simple and uniform scheme of substitution of geometric terms. Among the simpler examples are the following pairs of propositions:

- (1) For every two distinct points, there is exactly one line which is incident with both.
- (1') For every two distinct lines, there is exactly one point which is incident with both.
- (2) For every three points that are not incident with the same line, there is exactly one plane with which they are all incident.
- (2') For every three planes that are not incident with the same line, there is exactly one point with which they are all incident.

(1') results from (1) by interchanging 'point' and 'line'. They illustrate a general principle - the principle of duality for the projective plane - that covers many more instances. Similarly, (2) and (2') illustrate the principle of duality for projective space. (Statements (1') and (2') are not true for ordinary or 'metric' geometry, of course, but only for geometries in which the usual elements are augmented by various ideal elements such as points and lines at infinity.)

In addition to these simple and basic dualities, there are many others, some quite striking. They include a duality linking theorems of the Riemannian geometry of the plane to theorems of the Euclidean geometry of the sphere under interchange of the terms 'straight line' and 'great circle', and a duality linking theorems concerning lines in Euclidean three-dimensional geometry to theorems concerning points in Euclidean space of four dimensions under replacement of 'line' with 'point'.

Hilbert, however, was more taken with a duality that had been observed to exist between geometry - specifically, the axioms of linear congruence in Euclidean geometry - and a wholly non-geometric subject - namely, the laws governing proportions of trait-couplings in mutations of certain varieties of fruit flies (Hilbert 1922/23: 84-6). This duality suggested to Hilbert the idea that certain non-logical 'forms' of thought might be so basic or useful to our thinking that they would belong to even its most widely disparate parts (intuitively speaking). He therefore described it as 'more wonderful' than anything imagined in even the 'boldest fantasy' (Hilbert 1930: 380). Others, too (see Weyl 1927: 26-7, 1944: 635), were moved by dualities between remote subject matters.

The importance of the dualities for understanding Hilbert's formalist view is twofold. First, they suggest that in an axiom system, the mathematical terms do not function as constants but rather as variables of some sort. This in turn suggests that axiom systems are not to be seen as designed to capture a single, favoured interpretation that we think of as 'the truth'. Rather, they are to be seen as structuring a number of different subject matters. They are not designed to serve as theories of a particular subject matter but as 'theory-forms' - 'empty frames' (Pasch 1915:

11), 'logical moulds', 'hypothetico-deductive frameworks' (Weyl 1927: 25-6) - which provide a logical framework for the description of a variety of different subject matters (Hilbert 1899b: 40-1; see Detlefsen 1993 for more).

2 Axiomatization and the ideal method

The second respect in which the dualities of nineteenth-century projective geometry shed light on Hilbert's formalism is in their reliance upon the use of so-called 'ideal' or 'imaginary' elements. These are elements that (need) have no intuitional or perceptual basis and whose sole justification is the simplifying or generalizing effects they have on our thinking about a given subject. Their use is well-illustrated by the duality between (1) and (1') above. This depends upon the use of 'points at infinity' to serve as the points of incidence of pairs of parallel lines. (Without such ideal points, the duality between (1) and (1') would have to be reduced to one in which 'exactly' were replaced by 'at most'.)

Hilbert embraced the use of ideal elements in axiomatic theorizing. Indeed, he essentially identified the axiomatic method with the method of ideal elements (1926: 383). Specifically, he took the conditions for justified use of the ideal method to be the same as those for justified use of the axiomatic method; namely, that it be consistent with the underlying contentual practice to which it is applied and that it bring simplicity or efficiency to the production of mathematical knowledge (1926: 370, 372-3).

In Hilbert's view, then, the justification of an axiomatic system depends upon more than its mere consistency. It depends as well upon its promotion of 'epistemic efficiency', which is what use of the method of ideal elements has to offer. The geometric dualities illustrate this point well; for every contentual proof, they yield not one but two theorems - one directly provided by the contentual proof, the other by application of the substitution scheme of the duality. One thus, roughly speaking, obtains two theorems for the 'price' (that is, the genuine proof) of one. The second, 'free' theorem will, of course, imply not only real propositions but also ideal or imaginary propositions (for example, the intersections of parallel lines at infinity). Not everything that it covers thus constitutes a gain in real knowledge. Nevertheless, a great part of what it covers does, and it produces this with superior ease or efficiency.

Hilbert described the axiomatic or ideal method as the expression of an important intellectual freedom - namely, the freedom to create and use imaginary elements. This freedom could be taken as far as one wanted to take it provided that it did not clash with an associated body of contentual practice and that it promoted simplicity or efficiency in the production of real knowledge (1900: 439-40, 1905: 135-6, 1926: 372, 379). Dedekind also spoke of the 'freedom' of the axiomatic method. However, he understood it in a sense importantly different from Hilbert's - namely, as a freedom that contrasts axiomatic thinking not with contentual thinking, but with genetic thinking. Dedekind observed that genetically constructed objects (for example, the finite cardinals constructed as set-theoretic objects) inevitably take on features (for example, the infinity of the individual cardinals thus constructed) that are not relevant to their mathematical functioning. In Dedekind's view, the axiomatic method provides an alternative by giving us the freedom simply to declare that a set of items has exactly the properties laid down for them in a given set of axioms. Items introduced in this way are 'free creation(s) of the human intellect' (1888: §73) which retain only the relevant features of their genetically constructed counterparts.

Axiomatic freedom according to Hilbert, on the other hand, was the freedom to introduce admittedly non-real, imaginary entities into our thinking when doing so increases efficiency (over against purely contentual reasoning) in the production of contentual knowledge (1926: 370-73, 379, 392). He thus assigned ideal elements a role similar to that which Kant assigned the so-called 'ideas of reason' in his critical philosophy. Kant saw ideas of reason as non-descriptive, regulative, conservative devices for the efficient development of the judgments of the understanding (*Critique of Pure Reason*, B85-6, 362, 385, 536-7, 825-7). We are free to use them so long as we do not confuse them with objective, descriptive laws applying to an external reality (which confusion leads to inconsistency). Hilbert saw ideal elements as a psychologically efficient (perhaps even indispensable) means of developing our real or contentual knowledge: we are free to use them so long as doing so does not lead to inconsistencies with contentual thinking (1905: 135-6; 1926: 383, 392; 1928: 471).

3 The formalism of the arithmetical period

This basic conception of the ideal or axiomatic method was a key ingredient of Hilbert's thinking in both the

foundations of geometry and the foundations of arithmetic.

Hilbert's work in the foundations of arithmetic - his so-called 'programme' - was, indeed, nothing other than a defence of a particular application of the ideal method. What distinguished it from his work in the foundations of geometry was that the ideal elements used had a peculiarly 'logical' character, and that the consistency demand for use of these ideal elements could not be satisfied by reducing it to that of some more basic application of the ideal method.

The ideal elements that Hilbert proposed to defend in arithmetic were not imaginary mathematical items (for example, points at infinity, ideal numbers and the like) but imaginary *logical* items - what he referred to as 'ideal propositions'. These are not genuine propositions. They have neither meaning nor truth-value in the ('absolute') sense in which contentual propositions do. They are, rather, imaginary constructions (*Ideenbildungen*) that function only as symbolic or 'algebraic' devices for securing a desirable logic - namely, classical logic - for arithmetical reasoning. (See Hilbert 1922: 164-5; 1923: 179-80; 1926: 379-81; 1928: 467, 469-71; Hilbert and Bernays 1934: §1; see Hallett 1990 for a different interpretation.)

Examples that Hilbert gave of ideal propositions included, on the one hand, denials of true contentual generalizations (for example, ' $\neg\forall x(x + 1 = 1 + x)$ '), and, on the other, unbounded existential generalizations (for example, ' $\exists x(\text{Prime}(x) \ \& \ x > p)$ ', where p is some selected natural number) which follow classically from bounded, contentual existential generalizations (for example, ' $\exists x(\text{Prime}(x) \ \& \ x > p \ \& \ x < p!)$ '). The former are 'ideal' because they do not bound the search for a counter-example to the generalization denied; the latter because they do not bound the search for a witness of the type they assert to exist. However, both help to preserve classical logic as the logic of arithmetical thinking; the first by protecting the law of the excluded middle for true contentual generalizations, the second by protecting the classically valid inference-form from ' $\exists x(\phi x \ \& \ \psi x)$ ' to ' $\exists x(\phi x)$ ', where the former is a contentual truth.

Hilbert's claim is neither that there is no logic of finitary or contentual reasoning nor that, though there is, we cannot know it. Rather, it is that even if such a logic were to be given, we would not want and, perhaps, psychologically speaking, would not be able to use it. For even if given, it would diverge from that logic according to which we most naturally and efficiently reason towards truth in contentual arithmetic (1926: 379-80, 1928: 470-2).

Hilbert thus introduced ideal propositions as purely formal or algebraic instruments for preserving (on a formal level) the desirable patterns of classical logical inference. But this purely formal conception of the ideal propositions induced a similarly formal character for arithmetic as a whole (1926: 376-83, 1928: 467-70; Hilbert and Bernays 1934: §1) - the logical terms as well as the arithmetical terms. The logical terms occur in purely formal devices (namely, ideal propositions) and must therefore themselves be treated from a purely formal point of view. Hilbert's formal conception of ideal propositions thus drove his formalization of arithmetic beyond where his formalization of geometry had gone - namely, to the inclusion of logical as well as non-logical terms (see Detlefsen 1993).

4 The consistency of arithmetic

As we will now see, this radical formalization of ideal arithmetic is what motivated Hilbert's belief in the possibility of a direct or finitary proof of its consistency.

Reasoning from the radical formalization of ideal arithmetic, Hilbert offered a purely syntactic formulation of its consistency problem; namely, to show the nonexistence of a formal proof whose endformula is ' $0 \neq 0$ ', or some similarly contentually refutable formula. In like spirit, he believed that there ought to be a purely finitary, contentual resolution of this problem. The problem is to search for a certain type of 'formal object'. The negative resolution of such a problem would, it would seem, lie as much within the province of finitary reasoning as, say, the task of proving the non-existence of two numerals a and b satisfying the equation ' $a^2 = 2b^2$ ', a task which is finitarily manageable (1926: 383, 1928: 471).

Hilbert therefore undertook to give such a proof. It would, he believed, fill the final gap in the foundational enterprise by founding elementary ideal arithmetic. It would do so because it would be founded upon the 'absolute' truths of finitary judgment - judgments which neither need nor admit of any more basic justification

(1926: 376-7, 1922: 163; Hilbert and Bernays 1934: 20-1), and because the consistency of the other theories of ideal mathematics (most particularly those in analysis and geometry) can be reduced to that of ideal arithmetic (together with some basic set-theoretic apparatus).

This 'programme' of Hilbert's came under heavy attack in 1930-1 with the discovery of Gödel's incompleteness theorems. The second of these (G2), in particular, was important in this connection (see Gödel's theorems §6; Proof theory §2). It showed, for any formal axiomatic system T containing a fragment of arithmetic, how to construct a sentence Con_T of the language of T which seemingly 'expresses' the claim that T is consistent. It then went on to show that Con_T cannot be proved in T so long as T is consistent. If, therefore, T is rich enough to express all finitary proofs, it would follow that Con_T cannot be proved finitarily. From this it has generally been inferred that a finitary consistency proof is not possible for virtually any interesting portion of classical mathematics. Hilbert's programme therefore fails.

Whether such reasoning is correct depends, of course, upon whether Con_T really does 'express' the consistency of T in the appropriate sense. This, therefore, is one issue that must be resolved before the argument against Hilbert's programme from G2 can be conclusively evaluated. However, even if it is granted that Con_T does express the consistency of T , a further serious question arises: namely, whether the properties of Con_T that make for its unprovability in T (for example, the so-called derivability conditions of Hilbert-Bernays-Löb, or the RE condition of Feferman - see Gödel's theorems §5) are also properties that are indispensable to its ability to 'express' the consistency of T . Only if they are can the unprovability in T of Con_T be taken as showing the finitary unprovability of T 's consistency.

This deep and difficult problem has received little attention in the literature on Gödel's theorems and Hilbert's programme. But there are exceptions - see Detlefsen (1986: ch. 4) for a critical discussion of them. Detlefsen (1986, 1990) discusses what an adequate solution to the problem would look like and establishes the main features of two different types of negative solutions to it. Auerbach (1992) continues this discussion. In the end, the problem leads to fundamental questions concerning the nature of formal systems and to the perhaps even more fundamental question of what should be regarded as the basic unit of epistemic commitment in mathematics (see Detlefsen 1990 for further discussion).

Gödel's first incompleteness theorem (G1) has also been used against Hilbert's programme (see Kreisel 1976; Smoryński 1977, 1985, 1988; Prawitz 1981; Simpson 1988). The usual argument here begins with the basic premise that ideal arithmetic should be required to be a conservative extension of finitary or real arithmetic. G1 is then used to argue that this requirement cannot be met. The reasoning proceeds as follows: G_T (the undecidable sentence for a given ideal system of arithmetic T) can be proved by ideal but not by finitary methods; the ideal system in which G_T is provable is therefore not a conservative extension of finitary arithmetic; hence, by the basic premise of the argument, Hilbert's programme cannot be carried out. Detlefsen (1990) critically examines this argument, questioning, in particular, whether Hilbert's programme is rightly seen as requiring that ideal arithmetic be a conservative extension of real arithmetic. See also Webb (1980) for an argument that G1 actually confirms a formalist position like Hilbert's.

The Gödelian challenge to Hilbert's programme has also called forth other types of responses (see Proof theory §§3-4) - responses that concede the legitimacy of the arguments from G1 and G2, but which maintain that Hilbert's programme can none the less survive in a suitably modified form. For the most part, these arguments have taken one of two basic forms. In the first, the key element is weakening the evidential standards for consistency proofs to something less restrictive than Hilbert's original finitary demand. Some in this group (for example, Gentzen, Ackermann and Gödel) have argued that there are types of evidence having the same basic epistemic virtues as finitary evidence that extend well beyond the (traditionally identified) boundaries of finitary reasoning. Others (see Kreisel 1958; Feferman 1988; Sieg 1988) have argued not so much for a reconsideration of what should count as admissible evidence in a consistency proof as for a liberalization of our concept of what a consistency proof should do. They argue that the basic obligation of a consistency proof is to bring about some epistemic gain, and that there are epistemically gainful means of proving the consistency of ideal arithmetic that are not finitary. They therefore urge the replacement of the finitary/non-finitary distinction by a more refined hierarchical scheme that distinguishes grades of constructivity. Doing so, they believe, allows a significant partial realization of Hilbert's programme (see Proof theory §§2-4).

The other main type of response to Gödel's challenge is that represented by the so-called 'reverse mathematics' school of Friedman, Simpson and others. Instead of arguing for an extension of the types of reasoning to be allowed in consistency proofs, it aims at reducing the strength of the systems of ideal reasoning whose consistency needs proving. It begins by identifying more exactly those parts of traditional ideal mathematics that are truly indispensable to the distinctive achievements of classical mathematics. It then seeks axiomatizations that are 'equivalent to' these core parts and that therefore eliminate the extraneous elements of the usual axiomatizations. The idea is that the consistency of these diminished bodies of ideal reasoning might be susceptible to proof of a type to which that of their stronger standard counterparts is not. So far, significant success has been achieved along these lines (see [Simpson 1988](#); [Proof theory §4](#)).

See also: [Arithmetic, philosophical issues in](#); [Constructivism in mathematics](#); [Geometry, philosophical issues in](#); [Intuitionism](#); [Kronecker, L.](#); [Logical and mathematical terms, glossary of](#); [Mathematics, foundations of](#)

MICHAEL DETLEFSEN

References and further reading

- Auerbach, D.** (1985) 'Intensionality and the Gödel Theorems', *Philosophical Studies* 48: 337-51. (Discussion of the intensional aspects of G2.)
- Auerbach, D.** (1992) 'How to Say Things with Formalisms', in M. Detlefsen (ed.) *Proof, Logic and Formalization*, London: Routledge. (Detailed defence of the view that G2 refutes Hilbert's programme.)
- Bernays, P.** (1930) 'Die Philosophie der Mathematik und die Hilbertsche Beweistheorie' ('The Philosophy of Mathematics and Hilbert's Proof Theory'), *Blätter für Deutsche Philosophie* 3, 4: 326-67; repr. in *Abhandlungen zur Philosophie der Mathematik*, Darmstadt: Wissenschaftliche Buchgesellschaft, 1976. (Philosophical discussion of Hilbert's proof theory.)
- Bernays, P.** (1935) 'Hilberts Untersuchungen über die Grundlagen der Arithmetik' ('Hilbert's Investigations in the Foundations of Arithmetic'), in D. Hilbert, *Gesammelte Abhandlungen*, vol. 3, Berlin: Springer. (Philosophical and technical discussion of Hilbert's ideas in the foundations of arithmetic by his student and closest collaborator on foundational matters.)
- Dedekind, R.** (1888) *Was sind und was sollen die Zahlen?*, Braunschweig: Vieweg; trans. W.W. Beman (1901), 'The Nature and Meaning of Numbers', in *Essays on the Theory of Numbers*, New York: Dover, 1963. (§73 includes an interesting statement of Dedekind's philosophical conception of the axiomatic method and its relationship to the older, genetic method.)
- Detlefsen, M.** (1986) *Hilbert's Program: An Essay on Mathematical Instrumentalism*, Boston, MA, and Dordrecht: Reidel. (Detailed examination of the arguments from Gödel's incompleteness theorems against Hilbert's programme. Focuses primarily on the arguments from G2 and the question of what is required of formulas that are to express consistency.)
- Detlefsen, M.** (1990) 'On an Alleged Refutation of Hilbert's Program using Gödel's First Incompleteness Theorem', *Journal of Philosophical Logic* 19: 343-77. (Critical appraisal of the arguments against Hilbert's programme from G1.)
- Detlefsen, M.** (1993) 'Hilbert's Formalism', *Revue internationale de philosophie* 47: 285-304. (Attempt to clarify Hilbert's 'formalist' conception of mathematics and its relation to Kant's views.)
- Feferman, S.** (1988) 'Hilbert's Program Relativized: Proof-Theoretical and Foundational Reductions', *Journal of Symbolic Logic* 53: 364-84. (Useful introduction to the author's conception of a workable and interesting modification of Hilbert's original programme.)
- Hallett, M.** (1990) 'Physicalism, Reductionism and Hilbert', in A. Irvine (ed.) *Physicalism in Mathematics*, Dordrecht: Kluwer. (Argues for a non-instrumentalist reading of Hilbert.)
- Hallett, M.** (1994) 'Hilbert's Axiomatic Method and the Laws of Thought', in A. George (ed.) *Mathematics and Mind*, New York: Oxford University Press. (Discussion of Hilbert's conception of the axiomatic method.)
- Hallett, M.** (1995) 'Hilbert and Logic', in M. Marion and R. Cohen (eds) *Quebec Studies in the Philosophy of Science*, vol. 1, Dordrecht: Kluwer. (Account of the relationship between Hilbert's conception of logic and his views in the foundations of arithmetic.)
- Hilbert, D.** (1899a) *Grundlagen der Geometrie*, Leipzig: Teubner, 7th edn, 1930; 2nd edn trans. L. Unger and P. Bernays, *Foundations of Geometry*, La Salle, IL: Open Court, 1971. (Influential treatment of geometry from the widely held formalist perspective of the late nineteenth century.)
- Hilbert, D.** (1899b) letters to Frege of 12 and 29 December 1899, trans. H. Kaal, in G. Gabriel *et al.* (eds) *Gottlob Frege: Philosophical and Mathematical Correspondence*, Chicago, IL: University of Chicago Press, 1980. (Basic statement of Hilbert's formalist view of theories as logical frameworks of concepts/relations.)

- Hilbert, D.** (1900) 'Mathematische Probleme. Vortrag, gehalten auf dem internationalen Mathematiker-Kongreß zu Paris 1900', *Nachrichten von der königlichen Gesellschaft der Wissenschaften zu Göttingen, mathematisch-physikalische Klasse* 253-97; trans. 'Mathematical Problems', *Bulletin of the American Mathematical Society* 8: 437-79, 1902. (Includes a number of interesting remarks concerning the nature of mathematical truth and the axiomatic method.)
- Hilbert, D.** (1905) 'Über die Grundlagen der Logik und der Arithmetik', in A. Krazer (ed.) *Verhandlungen des dritten internationalen Mathematiker-Kongresses in Heidelberg vom 8. bis 13. August 1904*, Leipzig: Teubner, 174-85; trans. 'On the Foundations of Logic and Arithmetic', in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967, 130-8. (Early statement of Hilbert's developing views on the foundations of arithmetic. Perhaps most important for its statement of the 'creative principle' (135-6), which provides an important insight into his conception of the axiomatic method.)
- Hilbert, D.** (1918) 'Axiomatisches Denken' (Axiomatic Thinking), *Mathematische Annalen* 78: 405-15; repr. in *Gesammelte Abhandlungen*, vol. 3, Berlin: Springer, 1935. (Transitional essay in which Hilbert attempts to formulate a general conception of the axiomatic method in mathematics and science.)
- Hilbert, D.** (1922) 'Neubegründung der Mathematik. Erste Mitteilung' ('New Foundations for Mathematics: Part 1'), *Abhandlungen aus dem mathematischen Seminar der Hamburgischen Universität* 1: 157-77; repr. in *Gesammelte Abhandlungen*, vol. 3, Berlin: Springer, 1935. (Perhaps the clearest technical statement of Hilbert's programme for the foundations of arithmetic prior to the statement in **Hilbert and Bernays (1934)**.)
- Hilbert, D.** (1922/23) 'Wissen und mathematisches Denken. Vorlesungen 1922/23' ('Knowledge and Mathematical Thinking'), lecture notes taken by W. Ackermann; repr. Göttingen: Mathematisches Institut, 1988. (Largely philosophical discussion of Hilbert's foundational ideas.)
- Hilbert, D.** (1923) 'Die logischen Grundlagen der Mathematik' ('The Logical Foundations of Mathematics'), *Mathematische Annalen* 88: 151-65. (Includes Hilbert's distinction between the kind of truth - 'absolute truth' - that characterizes contentual thinking and the 'truth' that pertains to axiomatic or ideal thinking.)
- Hilbert, D.** (1926) 'Über das Unendliche', *Mathematische Annalen* 95: 161-90; trans. 'On the Infinite', in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic 1879-1931*, Cambridge, MA: Harvard University Press, 1967. (Perhaps the fullest and philosophically most revealing of Hilbert's writings on the foundations of arithmetic. Points out certain connections between Hilbert's and Kant's philosophies of mathematics.)
- Hilbert, D.** (1928) 'Die Grundlagen der Mathematik', *Abhandlungen aus dem mathematischen Seminar der Hamburgischen Universität* 6: 65-85; trans. 'The Foundations of Mathematics', in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967. (Mature statement of Hilbert's views regarding the foundations of arithmetic. Includes an important philosophical characterization of the central philosophical motive of his proof theory (page 475).)
- Hilbert, D.** (1930) 'Naturerkennen und Logik' ('Logic and the Knowledge of Nature'), *Die Naturwissenschaften* 18: 959-63; repr. in *Gesammelte Abhandlungen*, vol. 3, Berlin: Springer, 1935. (Philosophically revealing statement of Hilbert's mature thinking on the foundations of arithmetic. Points out respects in which Hilbert's philosophical thinking represents a modification of Kant's.)
- Hilbert, D. and Bernays, P.** (1934) *Grundlagen der Mathematik (Foundations of Mathematics)*, vol. 1, Berlin: Springer, 2nd edn, 1968. (Classic work of mathematical logic; includes valuable statements of Hilbert's conception of the axiomatic method and finitary reasoning.)
- Kreisel, G.** (1958) 'Hilbert's Programme', *Dialectica* 12: 346-72; revised version in P. Benacerraf and H. Putnam (eds) *Philosophy of Mathematics: Selected Readings*, Cambridge: Cambridge University Press, 2nd edn, 1983. (Reflections of a leading contemporary foundational thinker on Hilbert's programme. Interesting postscript added in 1978.)
- Kreisel, G.** (1976) 'What Have we Learnt from Hilbert's Second Problem?', in *Proceedings of Symposia in Pure Mathematics*, vol. 28, Providence, RI: American Mathematical Society. (Discussion of what recent developments in mathematical logic (especially proof theory) have to teach us about Hilbert's programme.)
- Lambert, J.H.** (1766) 'Theorie der Parallellinien' (Theory of Parallel Lines), *Magazin für reine und angewandte Mathematik* 2: 137-64, 3: 325-58, 1786; repr. in P. Stäckel and F. Engel (eds) *Theorie der Parallellinien von Euclid bis auf Gauss*, Leipzig: Teubner, 1895.
- Nagel, E.** (1939) 'The Formation of Modern Conceptions of Formal Logic in the Development of Geometry', *Osiris* 7: 142-224; repr. in *Teleology Revisited and Other Essays in the Philosophy and History of Science*, New York: Columbia University Press, 1982, 195-259. (Partially traces the origins of certain elements of the type of formalist outlook represented in Hilbert's work in the foundations of geometry.)

- Pasch, M.** (1882) *Vorlesungen über neuere Geometrie (Lectures on Modern Geometry)*, Leipzig: Teubner; 2nd edn Berlin: Springer, 1926.(Influential modern statement of the view that rigour requires that the inferences in a proof be formalized.)
- Pasch, M.** (1915) *Mathematik und Logik*, Leipzig: Engelmann, 2nd edn, 1924.(Further statements of Pasch's views regarding the formalization of inference in mathematical proof.)
- Peckhaus, V.** (1990) *Hilbertprogramm und kritische Philosophie: das Göttinger Modell interdisziplinärer Zusammenhänge zwischen Mathematik und Philosophie (Hilbert's Programme and Critical Philosophy: The Göttingen Model of Interdisciplinary Cooperation between Mathematics and Philosophy)*, Göttingen: Vandenhoeck & Ruprecht.(Historical discussion of the influence of the philosophers of the Friesian school at Göttingen on Hilbert's foundational thought.)
- Prawitz, D.** (1981) 'Philosophical Aspects of Proof Theory', in G. Fløistad (ed.) *Contemporary Philosophy: A New Survey*, vol. 1, The Hague: Martinus Nijhoff.(Useful discussion of the bearing of Gödel's theorems on Hilbert's programme.)
- Schlick, M.** (1925) *Allgemeine Erkenntnislehre*, Berlin: Springer; trans. A. Blumberg, *General Theory of Knowledge*, Vienna: Springer, 1974.(Chapter 1, §7 includes useful discussion of the nature of axiomatization.)
- Sieg, W.** (1988) 'Hilbert's Program Sixty Years Later', *Journal of Symbolic Logic* 53: 338-48. (Discussion of the progress and changes in Hilbert's programme since its formulation in the 1920s.)
- Sieg, W.** (1990) 'Relative Consistency and Accessible Domains', *Synthese* 84: 259-97.(Examination of the connections between Hilbert's foundational ideas and those of Dedekind and others.)
- Simpson, S.** (1988) 'Partial Realization of Hilbert's Program', *Journal of Symbolic Logic* 53: 349-63. (Survey of the progress made towards realizing the aims of Hilbert's programme via the approach of 'reverse' mathematics.)
- Smorynski, C.** (1977) 'The Incompleteness Theorems', in J. Barwise (ed.) *Handbook of Mathematical Logic*, Amsterdam: North Holland.(Survey of some of the technical material pertaining to Gödel's theorems.)
- Smorynski, C.** (1985) *Self-Reference and Modal Logic*, New York: Springer.(Includes a clear statement of an argument that G1 destroys Hilbert's programme.)
- Smorynski, C.** (1988) 'Hilbert's Programme', *Centrum voor Wiskunde en Informatica Quarterly* 1 (4): 3-59. (Fullest presentation of Smorynski's argument that G1 refutes Hilbert's programme.)
- Tait, W.W.** (1981) 'Finitism', *Journal of Symbolic Logic* 78 (9): 524-46.(Philosophical discussion of the nature of finitary reasoning.)
- Toepell, M.-M.** (1986) *Über die Entstehung von David Hilberts Grundlagen der Geometrie (On the Development of David Hilbert's Foundations of Geometry)*, Göttingen: Vandenhoeck & Ruprecht.(Historical treatment of various of Hilbert's writings and lectures on the foundations of geometry.)
- Torretti, R.** (1978) *Philosophy of Geometry from Riemann to Poincaré*, Dordrecht: Reidel.(A survey of philosophical ideas (and their origins) concerning geometry during the period named in the title.)
- Webb, J.** (1980) *Mechanism, Mentalism, and Metamathematics: An Essay on Finitism*, Boston, MA: Reidel. (Insightful discussion of various of Hilbert's philosophical ideas and the bearing of Gödel's theorems upon them.)
- Weyl, H.** (1927) 'Philosophie der Mathematik und Naturwissenschaft', in *Handbuch der Philosophie*, Munich: Oldenbourg; revised and expanded trans. *Philosophy of Mathematics and Natural Science*, Princeton, NJ: Princeton University Press, 1949.(Excellent introduction to foundational concerns in mathematics and the natural sciences by one of the most profound thinkers of the twentieth century.)
- Weyl, H.** (1944) 'David Hilbert and his Mathematical Work', *Bulletin of the American Mathematical Society* 50: 612-54.(Survey of Hilbert's work including comments on his foundational ideas.)

Identity

Anything whatsoever has the relation of identity to itself, and to nothing else. Things are identical if they are one thing, not two. We can refute the claim that they are identical if we can find a property of one that is not simultaneously a property of the other. The concept of identity is fundamental to logic. Without it, counting would be impossible, for we could not distinguish in principle between counting one thing twice and counting two different things. When we have acquired the concept, it can still be difficult to make this distinction in practice. Misjudgments of identity are possible because one thing can be presented in many guises.

Identity judgments often involve assumptions about the nature of things. The identity of the present mature tree with the past sapling implies persistence through change. The non-identity of the actual child of one couple with the hypothetical child of a different couple is implied by the claim that ancestry is an essential property. Knowledge of what directions are involves knowledge that parallel lines have identical directions. Many controversies over identity concern the nature of the things in question. Others concern challenges to the orthodox conception just sketched of identity itself.

1 Exposition of a popular view

Identity is the relation that, necessarily, each thing has to itself and to nothing else. Thus Constantinople has the identity relation to Istanbul because Constantinople is Istanbul, the very same city. This relation is often called ‘numerical’ identity, to distinguish it from ‘qualitative’ identity, the relation of exact similarity. Although ‘identical’ twins might be qualitatively identical, they are not numerically identical, for there are two of them, not one. The formula ‘ $x = y$ ’ says that x and y are (numerically) identical.

Identity is governed by two basic logical principles, reflexivity and Leibniz’s Law (see [Leibniz, G.W. §11](#)). To say that the identity relation is reflexive is to say that for each thing x , $x = x$. Leibniz’s Law says that if $x = y$ then whatever is true of x is also true of y ; it is used in arguments such as ‘Jack the Ripper was in Whitechapel last night, the Prince of Wales was not in Whitechapel last night; therefore Jack the Ripper is not the Prince of Wales’. [Leibniz \(1704\)](#) was not the first to formulate this law, which was known in antiquity. It is sometimes called the ‘indiscernibility of identicals’, not to be confused with Leibniz’s principle of the ‘identity of indiscernibles’, a kind of converse, which says (implausibly) that qualitative identity implies numerical identity (see [Identity of indiscernibles](#)). Reflexivity and Leibniz’s Law characterize identity uniquely: it is provably impossible for two non-equivalent relations to satisfy both principles. The two principles also entail that identity is symmetric (if $x = y$ then $y = x$) and transitive (if $x = y$ and $y = z$ then $x = z$).

If $a = b$, Leibniz’s Law says that whatever is true of a is true of b . However, this permits the replacement of ‘ a ’ by ‘ b ’ only in contexts in which the expressions merely specify which thing is being talked about. For example, it is invalid to argue from ‘Jocasta = the mother of Oedipus’ and ‘Oedipus knows that he married Jocasta’ to ‘Oedipus knows that he married the mother of Oedipus’, for here ‘the mother of Oedipus’ does not merely specify a person; it specifies the description under which Oedipus is said to know that he married her (see [Propositional attitudes](#)). Looking at photographs of a mature tree and a sapling, one cannot use Leibniz’s Law to refute the hypothesis ‘The mature tree = the sapling’ on the grounds that ‘The mature tree is tall’ is true and ‘The sapling is tall’ is false. The noun phrases ‘the mature tree’ and ‘the sapling’ do not merely specify trees; they indicate the times with respect to which tallness is predicated. A correct understanding of Leibniz’s Law is needed if identity through change is not to seem contradictory.

A genuine consequence of Leibniz’s Law is the necessity of identity: things that are in fact identical could not have been distinct (see [Modal logic, philosophical issues in §4](#)). For suppose that $x = y$. Since x could not have been distinct from itself, x could not have been distinct from y . Thus x and y cannot be contingently identical: they must be identical in all circumstances. A more complex argument concludes that they cannot be contingently distinct. They are either necessarily identical or necessarily distinct. Consequently, they are either always identical or always distinct. For example, if my headache is identical with an event in my brain, then that headache has to be that event; neither could exist without the other. Of course, a sentence such as ‘Rome = the capital of Italy’ is contingently true, but that is because the description ‘the capital of Italy’ need not have specified the city it actually specifies. The example is consistent with the necessity of identity, however, for it does not imply that one

city could have been each of two.

An argument like that for the necessity of identity can be used to refute the idea that questions of identity need have no right answer. For example, it is sometimes held to be indeterminate whether a given mass of rock and ice is Everest (see [Vagueness §1](#)). However, it is determinate whether Everest is Everest. If Leibniz's Law and ordinary logic apply in this context, it follows that Everest is distinct from that mass, and the identity question has a right answer - a negative one. Thus the hypothesis of indeterminacy contradicts itself.

Statements of identity in natural languages often include a noun answering the question 'same what?', for example, 'Istanbul is the same city as Constantinople'. Since identity is uniquely characterized by its logic, the role of the noun is not to disambiguate 'same'; ' a is the same F as b ' is equivalent to ' a is an F and $a = b$ ' (from which ' b is an F ' follows by Leibniz's Law). Identity is not defined kind by kind. Rather, we use the pre-given notion of identity in defining kinds. For example, we explain the difference between rivers and collections of water molecules by saying that the same river contains different collections of water molecules at different times.

A 'criterion of identity' for a kind is a necessary and sufficient condition for members of the kind to be identical. Frege's criterion of identity for directions is that the directions of lines are identical if and only if the lines are parallel (1884). The criterion of identity for numbers is that the number of F things is the number of G things if and only if there is a one-to-one correlation between the F things and the G things. In such cases, members of the kind can be presented in various guises: directions as the directions of various lines; numbers as the numbers of various pluralities. The criterion of identity states the condition for two guises to be guises of one member of the kind. Without a grasp of these conditions, one would not know what directions or numbers were.

A criterion of identity for a spatiotemporal kind is expected to give the condition for a member of the kind at one place and time to be identical with a member of the kind at another place and time: for example, it may have to follow a continuous trajectory between these space-time points. An account may also be required of an object's identity across hypothetical circumstances. For example, one may hold that a member of the kind could not have originated at different places and times in different possible worlds. Spatiotemporal objects can be exactly similar without being identical, for they can originate at different times or places. In contrast, purely abstract objects cannot be exactly similar without being identical.

2 Alternatives

Every aspect of the preceding view has been questioned. Although it is popular, the philosophers who accept it all may be in a minority.

That identity is a relation has been denied, both on metaphysical grounds (it cannot relate two things) and on grammatical grounds (some assimilate 'is' in 'Constantinople is Istanbul' to the 'is' of predication as in 'Constantinople is crowded'). However, the logic of 'Čequals;' does single out a unique class of ordered pairs of objects to which it applies, which suffices to make identity a relation in some minimal sense.

The contrast between identity and indiscernibility has been challenged, on the grounds that indiscernibles satisfy the same descriptions and since a satisfies the description 'identical with a ', so does anything indiscernible from a . The obvious reply is that indiscernibles merely satisfy the same *intrinsic* descriptions, but it is hard to explain what 'intrinsicness' is.

Many applications of Leibniz's Law are problematic. Some deny its applicability to contexts that treat non-actual possibilities or even non-present times, thus excluding the derivation of the necessity or permanence of identity. The intention is, for example, to permit a pot to be contingently identical with the clay of which it is made, or the clay to be temporarily identical with the pot. One question is whether the envisaged restrictions on Leibniz's Law are *ad hoc*. They would not be if so-called identity between objects in different possible worlds or at different times could not be taken at face value but was somehow reducible to relations of qualitative similarity among counterpart objects each of which was confined to a single possible world or time. Such views have been supported by appeal to the difficulty of specifying what is essential to an object, for example, to what extent a particular ship could originally have been made from different planks of wood, or how much it could change without ceasing to exist. However, the proposed reductions are both complex and hard to reconcile with ordinary assumptions about the nature of everyday objects.

Identity has also been regarded as sometimes indeterminate. Although it is usually conceded that if things are identical then it is determinate whether they are identical, in some nonstandard logics it does not follow that if it is indeterminate whether they are identical then they are not identical. Such logics postulate an intermediate status that propositions can have between truth and falsity. One challenge to this view is to explain what it means for a proposition to be not true without being false.

Yet another nonstandard view is the doctrine that identity is always relative to an answer to the question ‘same what?’. For example, if *a* and *b* are copies of *Middlemarch*, this view disputes the inference from ‘*a* is the same edition as *b*’ and ‘*a* is a copy’ to ‘*a* is the same copy as *b*’. The relation of being the same edition will not satisfy Leibniz’s Law unrestrictedly, but the view needs to show that no other relation (for example, being the same copy) could satisfy Leibniz’s Law unrestrictedly, for such a relation would be a case of non-relative identity.

Many disputes about criteria of identity for particular kinds of entity (see **Personal identity**) concern the nature of those entities, not of identity itself. But even the concept of a criterion of identity is itself problematic. For example, can one give an adequate criterion of identity for events by saying that events are identical if and only if they have the same causes and effects? The problem is that the causes and effects of events include other events. The criterion is in some sense circular, but it is hard to state the requirement of non-circularity clearly. Even if that could be done, it is unclear why every kind of object should have a non-circular criterion of identity. There may be nothing more basic than identity to which identity could be reduced. The standard logic of identity demands no such reduction.

Perhaps identity will come to be regarded as a logical constant, no more problematic than, say, conjunction. If so, many of the issues mentioned above will remain difficult, but their difficulty will concern the nature of various kinds of object, not the relation of identity.

See also: **Continuants §1**; **Logical and mathematical terms, glossary of**

TIMOTHY WILLIAMSON

References and further reading

- Frege, G.** (1884) *Die Grundlagen der Arithmetik: eine logisch mathematische Untersuchung über den Begriff der Zahl*, trans. J.L. Austin, *Foundations of Arithmetic*, Oxford: Blackwell, 2nd edn, 1980, §§64-73.(On criteria of identity.)
- Kripke, S.A.** (1980) *Naming and Necessity*, Oxford: Blackwell.(A basic work on the connection between identity and necessity.)
- Leibniz, G.W.** (1704) *New Essays on Human Understanding*, trans. and ed. P. Remnant and J. Bennett, Cambridge: Cambridge University Press, 1981, bk 2, ch. 27.(Gives Leibniz’s view of identity and contrasts it with that of John Locke.)
- Noonan, H.W.** (ed.) (1993) *Identity*, Aldershot: Dartmouth.(Collects many relevant articles.)
- Wiggins, D.R.P.** (1980) *Sameness and Substance*, Oxford: Blackwell.(On Leibniz’s Law and the ‘same what?’ question.)
- Williamson, T.** (1990) *Identity and Discrimination*, Oxford: Blackwell.(On supposed cases of indeterminate identity.)

Identity of indiscernibles

The principle of the identity of indiscernibles states that objects which are alike in all respects are identical. It is sometimes called Leibniz's Law. This name is also frequently used for the converse principle, the indiscernibility of identicals, that objects which are identical are alike in all respects. Both principles together are sometimes taken to define the concept of identity. Unlike the indiscernibility of identicals, which is widely accepted as a logical truth, the identity of indiscernibles principle has frequently been doubted and rejected. The principle is susceptible of more precise formulation in a number of ways, some more dubitable than others.

1 Leibniz's principle

In his metaphysics, Leibniz frequently stated a principle that no two separate individual things could differ only numerically, that is, resemble one another in all their properties or 'intrinsic denominations' and yet not be one and the same thing. According to various scholastics, two things may differ merely numerically, meaning they could have the same nature, but be distinguished by their matter. Immaterial individuals such as angels must each have a distinct nature, and Leibniz, rejecting matter, extended this view to all individuals. Sometimes he justified it theologically: God never does anything without a reason, and if two things were exactly alike, he would have no reason to create one rather than the other. Sometimes Leibniz referred to empirical evidence, recounting the story of the nobleman he set vainly to searching a castle garden to find two leaves exactly alike, though of course the mere failure to find a counter-instance on a restricted search is no serious test. Leibniz used his principle to argue against empty space and time.

In his logical writings Leibniz defined a relation of sameness between terms as follows: 'Terms which can be substituted for one another whenever we please without changing the truth of any statement (*salva veritate*), are the same (*eadem*) or coincident (*coincidentia*)'. When terms are coincident they denote the same thing or things. Leibniz's definition covers general as well as singular terms, as his triangle/trilateral example makes clear. It would perhaps be better to say that the things are the same when their names are coincident, but Leibniz's meaning is clear.

Leibniz's definition of identity was adopted by Frege, who restricted it to singular terms. In second-order predicate logic, singular identity is often defined as follows:

$$a = b \equiv_{\text{df.}} \forall F(F(a) \equiv F(b)).$$

The left-to-right implication is uncontroversial, but needs careful formulation to exclude non-extensional contexts. For example, in 'John believes that x defeated Mark Antony', substituting the names 'Octavian' and 'Augustus' for x may yield different truth-values if John does not realize that Augustus and Octavian are the same person, yet his ignorance does not impugn their identity (see **Propositional attitudes**). F is taken to range over properties or predicate extensions determined by extensional sentential contexts only. It is thus important to distinguish the substitution principle, which is about expressions, and requires such restrictions, from the indiscernibility principle, which is about the properties of individuals (see **Identity §1**).

It is the right-to-left implication that is the modern symbolic form of Leibniz's identity of indiscernibles, and this is the half of the definition that has attracted most criticism. There is a version of the principle which is trivially true. Assume $\forall F(F(a) \equiv F(b))$. Substituting ' $a = x$ ' for ' $F(x)$ ' we have $(a = a) \equiv (a = b)$, and since it is logically true that $a = a$, it follows that $a = b$.

Leibniz hardly intended his principle to be understood thus: had it been a logical triviality, he would not have laboured it; indeed he did not consider it a logical truth. This version of the principle is not in doubt. To rescue it from triviality and make it fit to define identity, we must screen off identity and all properties involving it from consideration. Call properties that pass the supposed screening-off test 'material' properties. The most obvious material properties are qualities, so a more substantive version says that no two distinct individuals have all their qualities in common. This is closer to Leibniz's intentions, but readily refuted. For very simple and numerous individuals such as fundamental particles the qualities attributable to them are few, and it is unlikely that two such particles are never exactly alike with respect to the few qualities they do have.

Even allowing that a particle has a history and may change its qualities, we cannot exclude the possibility of two particles having exactly similar qualitative histories. Even if two things never in fact resemble one another exactly, it is not logically impossible that two distinct things be exactly alike. So the qualitative version of the principle seems clearly false.

2 Temporal and modal versions

Evidently there are several indiscernibility principles, differing in plausibility. The question becomes: which one is neither so weak that it is trivial nor so strong that it is false, but just strong enough to characterize identity?

Two objects may be ‘temporarily’ indiscernible and yet be distinct because they are discernible at some other time. A golden cup and the gold out of which it is made occupy the same place, have the same weight, thermal conductivity, stand in the same spatial and causal relations to other contemporaneous things, and so on. But the gold existed before the cup was fashioned out of it.

Suppose two objects came into being simultaneously and have been always indiscernible to date. Are they identical? Not if their histories diverge later. If a cat C and the mereological sum of its body and tail $B + T$ have been indiscernible until now, but later the cat loses its tail, then the histories of C and $B + T$ diverge, for C loses T as a part but $B + T$ does not (it either becomes scattered or ceases to exist), so C was not identical with $B + T$: they were merely indiscernible prior to the loss of the tail. The property by which they differed was future-related, just as the history of the gold before it came to constitute the cup is past-related.

Neither temporary indiscernibility nor indiscernibility to date suffice for identity, but if individuals are ‘permanently’ indiscernible, that is, exist and have the same properties at the same times, are they identical? It seems not. For suppose C never loses its tail. C remains coincident and indiscernible from $B + T$ at all times it exists. But C , unlike $B + T$, could have lost T and carried on existing: cats, unlike body-tail sums, can exist without tails. So it seems that C is ‘modally’ discernible from $B + T$. (See [Essentialism §4.](#))

Does identity require indiscernibility in all modal properties? This seems too strong. For example, it is necessary to the first Roman Emperor that he have subjects, but not necessary to Octavian that he have subjects, although Octavian was the first Roman Emperor. This example is inconclusive, since ‘the first Roman Emperor’ is a definite description and the necessity is *de dicto*: it is not necessary *de re* to the Emperor, that man, that he have subjects (see [De re/de dicto §1](#); [Descriptions §2](#)). This may seem to rebound on the cat example: ‘ $B + T$ ’ is also a definite description, ‘the sum of B and T ’, so the fact that C might not have coincided with $B + T$ does not entail that they are distinct, since by the theory of descriptions ‘ $C = \text{the } A$ ’ just means

$\forall x(x \text{ is an } A \equiv (x = C))$.

But the cases differ, for while it is true of Octavian that he was an emperor, it is not obviously true of the cat that it is a sum of body and tail. The difference is again modal: no emperor is essentially an emperor, but sums are essentially sums, and C is not essentially a sum (see [Essentialism §4](#)).

This gives us a more precise modal version: if a and b are indiscernible with respect to all their non-modal properties and all their ‘essential’ properties then $a = b$. What then is an essential property? One which a must have if it exists. But being identical with a is such a property, so triviality is again a problem. Furthermore, deciding whether a property is essential or not may turn on deciding identity questions. While this indicates that matters of essence are closely connected with identity, we have to restrict attention to material properties again, and modality muddies considerations.

3 Relations and thought experiments

Once we admit relational properties, that is, consider how individuals stand in relation to others, we radically enlarge the range we quantify over and enhance what can be discerned. Though two electrons might be qualitatively indiscernible, if one is 1 metre away from a third particle while the other is not, they are distinguished by this relational property. The Thomistic view that matter individuates bodies which are alike in nature is also relational: $a \neq b$ because m is a ’s matter and not b ’s.

This suggests a more promising non-modal version: individuals which are ‘materially’ indiscernible, that is,

indiscernible with respect to all material properties (monadic and relational), are identical. Can there be two distinct materially indiscernible individuals? Take the two electrons mentioned above. If the third particle is a proton and the electron that is not 1 m from it is 1 m from no other proton, then the electrons differ in their material relations. If both were 1 m from qualitatively indiscernible protons the electrons would not thereby differ materially, for though this electron is 1 m from this proton and that electron 1 m from that proton, we can no longer use this information to discern them, since it turns on this proton's being distinct from that one.

To find numerically distinct but materially indiscernible individuals, philosophers have concocted thought experiments with symmetrical universes. Max Black (1952) envisaged a universe consisting solely of two qualitatively indistinguishable spheres with indiscernible histories. Each sphere and each part of a sphere is materially indiscernible from a numerically distinct sphere or part of a sphere. Except for symmetrical individuals, every part of the universe has a numerically distinct *doppelgänger*. Black's universe appears possible, so either the spheres would have to be merely formally distinct, violating the principle, or there would have to be some haecceitas or individualizer to distinguish the spheres, each being uniquely related to its own haecceitas. But to posit such unobservable things seems unacceptably metaphysical.

Those who uphold the principle of material indiscernibility argue by *modus tollens* that there can be no symmetrical universe, for the spheres would be one sphere, and since the sphere is spherically symmetrical it would 'collapse' to a line, whose points, being identical, would collapse to a point: the only symmetrical universe would be monistic. Since our universe is apparently not symmetrical, it is hard to know how to resolve this dispute.

4 Empirical counterexamples

Harder and more disturbing evidence against Leibniz's principle in almost all versions comes from physics. Two electrons in an atomic shell are discernible from one another while both are in it, having opposite spins, but if an electron enters the shell and an electron later leaves it, there is no way to answer the question whether the same electron left as previously arrived. It is not merely that we cannot observe the proceedings closely enough: there is seemingly no fact of the matter. But electrons are fermions and so must always be discernible at a point in time by their quantum states. The Leibnizian may therefore retreat to holding that identity for electrons is only synchronic, and that they are only momentary, though lone electrons seem to persist happily, only gregarious ones 'losing their identity'. The case is worse for photons, which are bosons and so can be temporarily indiscernible in all their material properties, as well as their position and relations. They can become superposed 'soup': a laser is just very homogeneous photon soup. Yet the intensity of a laser beam depends on how many photons are in it: coincident photons do not collapse into one. A Leibnizian can rescue indiscernibility here only by denying that photons are individuals: talk of photons must then be a physicist's dispensable *façon de parler*.

So although there is dispute about what the indiscernibility of identicals amounts to, and whether in this or that version it is true, we appear comfortable conceiving of things as individuals only to the extent that they are in some sense discernible.

See also: [Logical and mathematical terms, glossary of](#)

PETER SIMONS

References and further reading

- Black, M.** (1952) 'The Identity of Indiscernibles', *Mind* 61: 153-64. (A famous dialogue on the spheres example.)
- Brody, B.A.** (1980) *Identity and Essence*, Princeton, NJ: Princeton University Press. (Tenaciously defends and exploits the trivial version of Leibniz's principle.)
- Fraassen, B.C. van** (1991) *Quantum Mechanics: An Empiricist View*, Oxford: Clarendon Press. (See chapters 11, 12 for a logically astute philosopher's discussion of the 'problem of identical particles' in physics.)
- Leibniz, G.W.** (1956) *Philosophical Papers and Letters*, ed. L.E. Loemker, Dordrecht: Reidel, 1975. (Follow the index entries under 'Identity'.)
- Mates, B.** (1986) *The Philosophy of Leibniz: Metaphysics and Language*, New York: Oxford University Press. (Chapter 7, §3 discusses the identity of indiscernibles principle in Leibniz.)
- Wiggins, D.** (1980) *Sameness and Substance*, Oxford: Blackwell. (Chapter 2, §2 discusses (and rejects) the identity of indiscernibles principle amid a general theory of identity.)

Imperative logic

Imperatives lie at the heart of both practical and moral reasoning, yet they have been overshadowed by propositions and relegated by many philosophers to the status of exclamations. One reason for this is that a sentence's having literal meaning seems to require its having truth-conditions and 'Keep your promises!' appears to lack such conditions, just as 'Ouch!' does. One reductionist attempt to develop a logic of imperatives translates them into declaratives and construes inferential relations among the former in terms of inferential relations among the latter. Since no such reduction seems fully to capture the meaning of imperatives, others have expanded our notion of inference to include not just truth - but also satisfaction - preservation, according to which an imperative is satisfied just in case what it enjoins is brought about.

A logic capturing what is distinctive about imperatives may shed light on the question whether an 'ought' is derivable from an 'is'; and may elucidate the claim that morality is, or comprises, a system of hypothetical imperatives. Furthermore, instructions, which are often formulated as imperatives ('Take two tablets on an empty stomach!'), are crucial to the construction of plans of action. A proper understanding of imperatives and their inferential properties may thus also illuminate practical reasoning.

1 Jörgensen's dilemma and Ross's 'paradox'

Any attempt to codify the logical properties of imperatives must grapple with the following dilemma, first formulated by **Jörgensen (1938)**. On the one hand there appear to be cogent inferences involving imperatives, as in:

Keep your promises!
This is a promise of yours.
Keep this promise!

On the other hand, the only widely agreed-upon conception of good inference is in terms of truth-preservation: a set of sentences Γ implies a sentence ϕ just in case there is no possible situation in which all members of Γ are true and ϕ is false. Yet evidently this conception of cogency in terms of truth-preservation cannot be used to assess the validity of arguments utilizing imperatives, for imperatives are not, *prima facie*, either true or false. Hence either we deny, implausibly, that there are good inferences involving imperatives; or we grant it and then face the question, to what conception of inference are we appealing?

This challenge is to be distinguished from another, less urgent concern first formulated by **Ross (1944)** and echoed by **Williams (1963)**. According to the classical logic of truth functions, any proposition P implies $P \vee Q$, for any Q . Similarly, it would seem that even the most conservative imperative logic would countenance the inference

Post this letter!
Post this letter or burn it!

Demanding that the auditor either post the letter or burn it may be seen as granting permission to do either. But then if anyone committed to the premise is thereby committed to the conclusion above, it may seem that demanding that the auditor post the letter is implicitly granting permission to burn it. This reasoning vacillates on the concept of commitment. We may distinguish between commitment to the content of a sentence and commitment to performing the speech act of uttering that sentence with its characteristic illocutionary force. One who utters the premise of the above inference is not thereby committed to performing the speech act of demanding that the auditor either post the letter or burn it, just as one who asserts P is not thereby committed to asserting $P \vee Q$. (Asserting P merely commits one to the truth of $P \vee Q$.) Yet it is only the performance of the speech act of uttering an imperative that grants permission to the auditor to satisfy that sentence by satisfying any disjunct that it contains: commitment merely to the content of that sentence does not suffice to grant permission of any kind.

To others a logic of imperatives may seem trivially simple. It is common to distinguish between two elements in an imperative sentence, one of which is shared with indicative sentences and the other not: from both 'Shut the door!' and 'The door is shut' we can abstract a reference to the state of affairs of the door's being shut. The former

sentence enjoins that state of affairs while the latter asserts that it obtains. Following [Stenius \(1967\)](#), let us say that what is common to the two sentences is the ‘sentence-radical’, while what distinguishes them is their ‘modal element’. One might urge that the validity of arguments involving imperatives should be understood just in terms of sentence-radicals, since these require no modification in our ‘truth-preservation’ conception of validity. But this leaves us with no account of when and where modal elements can occur in valid arguments. It is unclear, for instance, how without *ad hoc* postulation we are to avoid countenancing:

Shut the door!
The door is shut.

Approaches to the logic of imperatives may be seen as grappling with one or other horn of Jørgensen’s dilemma.

2 Reductionist approaches

Reductionist approaches attempt to show that there is no good inference that crucially involves imperatives; at best, any inference that seems to involve imperatives can be replaced with an equivalent one comprising only indicatives, to which the truth-preservation conception of validity applies. To the objection that only propositions appear to be truth-bearers, reductionists might reply that the difference in grammatical mood between imperatives and declaratives is an idiosyncrasy of idiom; a mere reflection of the fact that these two types of sentence are typically used to perform different kinds of speech acts. The difference of mood need be of no more interest to the logician than is the difference between the active and passive voice in declarative propositions.

So with what propositions are imperatives to be identified? On one proposal, ‘Shut the door!’ is equivalent to ‘You will shut the door’. This equation is perhaps suggested by the fact that the latter type of sentence is often used to convey a command; if correct then imperative inference may be understood in terms of inferential relations among statements in the future tense (see [Tense and temporal logic](#)). Yet this identification falls foul of our ability to make predictions without formulating any kind of directive: a parole officer might say to a repeat offender, ‘You will steal again’, without enjoining him to crime. Others have contended that ‘Shut the door!’ is equivalent to ‘I want you to shut the door’; here imperative inference is a direct consequence of the logical properties of attitude locutions (see [Propositional attitude statements §§1-2](#)). But one can, and may have to, command something one does not want. Furthermore, recipes and instructions are often formulated with imperatives but appear not to be reports of anybody’s desires.

It has also been proposed that ‘Shut the door!’ is equivalent to ‘Either you are going to shut the door, or *X* will happen’, where *X* is something bad for the person being addressed. Yet a speaker might reasonably utter an imperative such as ‘Keep your promises!’ knowing full well that no harm will come the auditor if they fail to obey. Second, the theory implies that for any action *Y*, either ‘Do *Y*!’ or ‘Don’t do *Y*!’ is true. A distinct, ‘performative’ analysis of the imperative has it that ‘Shut the door!’ is equivalent to ‘I order you to shut the door’ (see [Performatives](#)). This proposal requires an imperative to be as multiply ambiguous as there are distinct illocutionary acts that can be performed with that sentence; on a given occasion of utterance the same sentence might mean, ‘I request that you shut the door’. This theory is thus committed to a multiplication of senses for what seem to be univocal sentences.

The most popular reduction may be traced to [Kant \(1785\)](#) and equates ‘Shut the door!’ with the ‘deontic’ proposition ‘You should (or ought to) shut the door’. On this approach imperative inference is a matter of inferential relations among deontic locutions (see [Deontic logic](#)). We may distinguish between two species of imperative - fiats and directives - only the latter of which are directed to a particular person or group of persons. Clearly, directives are the much more common form and for these it would seem that being directed to a particular individual is part of their content. This is perhaps why imperatives are so naturally issued with vocatives, as in ‘Mary, make sure John does his homework’. In contrast to an imperative, the truth of a deontic proposition does not, as such, lay obligations upon any one person rather than another. ‘It is obligatory that *A*’ does not discriminate among the various people who might be obliged to make it the case that *A*. This even holds for ‘John should see to it that *A*’: a speaker might want to lay responsibility on Mary for making John see to it that *A*, but will not be able to do so with just deontic propositions. A deontic reduction lacks the resources to capture the directedness of imperatives.

3 Non-reductionist approaches

Grappling with the other horn of Jørgensen's dilemma involves formulating a notion of inference to include things other than truth-bearers. Since validity as here characterized is a semantic notion, an account of valid reasoning with imperatives would require assigning semantic values to imperative sentences. One proposal stems from **Lemmon (1965)**, according to whom the basic semantic value of an imperative is to be that of its being 'in force' (or not). Thus when a commanding officer says, 'Clean the latrine!' to a soldier, the command is typically in force, whereas if the soldier were to say the same to the commanding officer this would typically not be the case. This suggests the following criterion of validity:

An argument is valid if and only if, given that all the directive premises are in force and all the declarative premises are true, then the conclusion is in force if it is a directive and true if it is a declarative.

This definition of validity sheds no light on the question whether, for instance, the being in force of 'Take all the boxes!' implies the being in force of 'Take this thing!' when the speaker is unaware that this thing is one of the boxes. Lemmon's proposal is thus a mere schema for a definition. A more explicit proposal flows from the work of **Hofstadter and McKinsey (1939)**, **Rescher (1966)** and **Sosa (1967)**. Say that an imperative is 'satisfied' just in case its sentence-radical (see above) is true. One might suggest that an argument is valid if and only if, whenever its imperative premises are satisfied and its indicative premises are true, then its conclusion is satisfied if imperative and true if indicative. This definition would explain the apparent validity of the inference with which §1 of this entry began. On the other hand, it countenances the inference from 'Shut the door!' to 'The door is shut'.

Sosa attempts a definition in terms of satisfaction-preservation for arguments consisting only of imperatives, some of which may be hypothetical in form. Let us say that imperatives such as 'If it rains, close the window!' are 'satisfied' if it rains and the window is closed; 'violated' if it rains and the window is not closed; and 'neutral' if it does not rain. (All imperatives can be construed as hypothetical if we imagine $P \vee \sim P$ as antecedent when none is otherwise specified.) We now say:

A directive argument is valid just in case (1) if its premises are satisfied then the conclusion is satisfied and (2) if the conclusion is violated then at least one of its premises is violated.

This definition can account for the validity of

If you read this book, come to see me!
Read this book!
Come to see me!

On the other hand, the definition renders invalid the apparently correct

If it rains, close the window!
If it rains and thunders, close the window!

A cognate definition given by Rescher founders on its inability to treat the relation 'is validly inferable from' as being transitive.

There is today no accepted account of imperative inference. Instead research focuses upon the semantics of imperatives while stopping short of formulating a calculus. Yet if these semantic questions can be answered it may be hoped that new proposals for understanding imperative inference may grow in the light the answers shed. Until that time reductionist approaches will continue to be attractive.

See also: **Emotive meaning**; **Logical and mathematical terms, glossary of**; **Pragmatics**; **Questions**; **Speech acts**
MITCHELL GREEN

References and further reading

- Austin, J.L.** (1962) *How to Do Things with Words*, Oxford: Clarendon Press. (A classic work in the theory of speech acts. Defends a version of the 'I order you to...' species of reduction.)
Beardsley, E. (1944) 'Imperative Sentences in Relation to Indicatives', *Philosophical Review* 53: 175-85. (Early proponent of non-reductionism.)

- Belnap, N.** (1989) 'Declaratives Are Not Enough', *Philosophical Studies* 59: 1-30. (Attacks the sentence-radical/modal element distinction as failing to capture the distinctive content of imperatives.)
- Bohnert, H.** (1945) 'On the Semiotic Status of Commands', *Philosophy of Science* 12: 302-15. (Argues for the reduction of imperatives to sentences of the form 'Either you conform or something bad will happen'.)
- Castañeda, H.-N.** (1975) *Thinking and Doing: The Philosophical Foundations of Institutions*, Dordrecht: Reidel. (Wide-ranging study of imperatives, particularly their semantics and their role in practical reasoning.)
- Chellas, B.** (1969) 'Imperatives', *Theoria* 37: 114-29. (Formulates a logic for imperatives, treating them as reducible to deontic propositions.)
- Frege, G.** (1892) 'Über Sinn und Bedeutung', *Zeitschrift für Philosophie und philosophische Kritik* 100: 25-50; trans. M. Black, 'On Sense and Reference', in *Translations from the Philosophical Writings of Gottlob Frege*, ed. P.T. Geach and M. Black, Oxford: Blackwell, 2nd edn, 1960. (Historically important formulation of a distinction between illocutionary force and semantic content.)
- Hamblin, C.L.** (1987) *Imperatives*, New York: Blackwell. (Best full-length treatment to date of the grammar, semantics and pragmatics of imperatives. Includes an extensive bibliography.)
- Hare, R.M.** (1952) *The Language of Morals*, Oxford: Clarendon Press. (Propounds the centrality of imperatives in ethical discourse. Distinguishes between phrastic and neustic.)
- Hofstadter, A. and McKinsey, J.C.C.** (1939) 'On the Logic of Imperatives', *Philosophy of Science* 6: 446-57. (Mentioned in §3. Early formulation of the notion of satisfaction as a distinctive semantic value for imperatives.)
- Huntley, M.** (1984) 'The Semantics of English Imperatives', *Linguistics and Philosophy* 7: 103-33. (Sophisticated defence of a semantic treatment of imperatives as differing from propositions in lacking an indexical element.)
- Jørgensen, J.** (1938) 'Imperatives and Logic', *Erkenntnis* 7: 288-98. (Mentioned in §1. Explains the central challenge facing attempts to formulate an imperative logic.)
- Kant, I.** (1785) *Grundlegung zur Metaphysik der Sitten*, trans. H.J. Paton, *Groundwork of the Metaphysics of Morals*, London: Harper Perennial, 1989. (Mentioned in §2. Defends a reduction of imperatives to deontic propositions.)
- Lemmon, E.J.** (1965) 'Deontic Logic and the Logic of Imperatives', *Logique et Analyse* 8: 39-71. (Mentioned in §3. Sketches a conception of imperative inference in terms of the concept of an imperative's being in force.)
- Lewis, D.K.** (1970) 'General Semantics', repr. in *Philosophical Papers*, vol. 1, Oxford: Oxford University Press, 1983. (A classic paper in the philosophy of language. Defends a version of the 'I order you to...' species of reduction.)
- Rescher, N.** (1966) *The Logic of Commands*, London: Routledge & Kegan Paul. (Mentioned in §3. Highly readable general treatment of imperative logic as well as an excellent introduction to the topic. Defends a notion of imperative inference in terms of satisfaction-preservation.)
- Ross, A.** (1944) 'Imperatives and Logic', *Philosophy of Science* 11: 30-46. (Mentioned in §1. Argues for the impossibility of a non-trivial imperative logic.)
- Searle, J.R. and Vanderveken, D.** (1985) *Foundations of Illocutionary Logic*, Cambridge: Cambridge University Press. (Develops some logical features of devices used to indicate illocutionary force and inferential relations among various types of speech act.)
- Sosa, E.** (1967) 'The Semantics of Imperatives', *American Philosophical Quarterly* 4: 57-64. (Mentioned in §3. Defends a notion of imperative inference in terms of satisfaction-preservation.)
- Stenius, E.** (1967) 'Mood and Language-Game', *Synthese* 17: 254-74. (Mentioned in §1. Defends a distinction between the sentence-radical and modal element.)
- Williams, B.A.O.** (1963) 'Imperative Inference', *Analysis*, supplement 23: 30-6. (Mentioned in §1. Following Ross, argues for the impossibility of an imperative logic.)

Indicative conditionals

Examples of indicative conditionals are *'If it rained, then the match was cancelled'* and *'If Alex plays, Carlton will win'*. The contrast is with subjunctive or counterfactual conditionals, such as *'If it had rained, then the match would have been cancelled'*, and categoricals, such as *'It will rain'*.

Despite the ease with which we use and understand indicative conditionals, the correct account of them has proved to be very difficult. Some say that *'If it rained, the match was cancelled'* is equivalent to *'Either it did not rain, or the match was cancelled'*. Some say that the sentence asserts that the result of 'adding' the supposition that it rained to the actual situation is to give a situation in which the match was cancelled. Some say that to assert that if it rained then the match was cancelled is to make a commitment to inferring that the match was cancelled should one learn that it rained. This last view is often combined with the view that indicative conditionals are not, strictly speaking, true or false; rather, they are more or less assertible or acceptable.

1 Indicative and material conditionals

In general an 'indicative conditional' has the form 'If A then C ', where A is called the antecedent and C the consequent. A central issue is the relationship between the truth-value of a conditional and the truth-values of its antecedent and consequent. This much is immediately plausible: if A is true and C is false, then the conditional is false. If I say 'If it rained, the match was cancelled', and what happened was that it rained but the match went ahead, then what I say is clearly false. There are three other possibilities: A and C are both true, A is false and C is true, and A and C are both false. There are a number of arguments designed to show that in each of these cases the conditional is true. Here is one. 'If A then C ' is intuitively equivalent to the disjunction 'Either not- A , or A and C '. (Instead of saying that if it rained the match was cancelled, I could have said 'Either it did not rain, or it did and the match was cancelled'.) But the latter is true in all three cases: in each, either the first disjunct ('not- A ') or the second disjunct (' A and C ') is true. A conditional that is false when its antecedent is true and its consequent false, and true in all other cases, is called a 'material conditional', and is symbolized ' $A \supset C$ ' (read ' A hook C '). ' $A \supset C$ ' is definitionally equivalent to 'Not- A or C '. Hence the argument's conclusion is that indicative conditionals are equivalent to material conditionals. This conclusion has the virtue of validating the two most obviously valid inferences governing conditionals - *modus ponens*: ' A , if A then C , therefore, C ', and *modus tollens*: 'Not- C , if A then C , therefore, not- A '.

Despite the appeal of the argument, there are serious problems for the equivalence thesis that indicative conditionals are material conditionals. It entails that any conditional with a false antecedent is true regardless of its consequent, that is, the account validates 'Not- A , therefore, if A then C '. This is implausible. 'If I live in London then I live in Scotland' strikes us as false (it is rather 'If I live in London then I do not live in Scotland' which is true), but because I do not live in London, it is true on the equivalence thesis. Also, it entails that any conditional with a true consequent is true: the account validates ' C , therefore, if A then C '. But 'If I live in London, then I live in Australia' strikes us as false even after we learn that I do in fact live in Australia. These two results are known as the paradoxes of material implication ('material implication' being the name of the relation between A and C when ' $A \supset C$ ' is true).

2 Possible worlds treatments

An obvious response to the paradoxes is to insist that more than the truth of ' $A \supset C$ ' is required in order for 'If A then C ' to be true; that the truth of ' $A \supset C$ ' is a necessary but insufficient condition. One way to do this is to require a connection between antecedent and consequent; to hold that part of what makes 'If it rains, the grass will grow' true is the connection between rain and grass growing. This blocks the paradoxes. However, we sometimes use conditionals precisely to deny that there is a connection between antecedent and consequent. A doctor who says that if you go to bed and take an aspirin you will get better in a week, whereas if you go to work it will take seven days, is saying that there is no connection between going to bed and getting better.

A more promising approach is to require that the material conditional not only be true the way things actually are, but be true in the closest possible worlds where the antecedent is true (see [Semantics, possible worlds](#)). Roughly, the account is: 'If A then C ' is true iff ' $A \supset C$ ' is true, that is, true the way things actually are *and also* true in

the closest possible worlds where A is true. Another way of saying the same thing is: ‘If A then C ’ is true iff the closest worlds in which A is true are worlds in which C is also true. (This is because any world where A and ‘ $A \supset C$ ’ are true is a world where C is true.) The appeal of this approach derives from the appeal of the idea that when we evaluate a conditional we look at the way things actually are and ‘in imagination’ add the antecedent and then see whether, with only the changes forced by adding the antecedent, the consequent comes out true.

This approach is very attractive for subjunctive or counterfactual conditionals (see [Counterfactual conditionals](#)) but faces a problem when applied to indicative conditionals. (Interestingly, one architect of the possible worlds approach to conditionals, David Lewis (1973), only ever intended it to apply to subjunctive conditionals. Robert Stalnaker, the other architect (1968), intended it to apply to both.) It construes indicative conditionals as being in part about the way things might have been but are not in fact. That seems exactly right for subjunctives. When I say that had I invested in Western Mining ten years ago, I would now be rich, I am saying how things are in some possible but regrettably non-actual world. Indeed, I might express this by saying that had I invested in Western Mining ten years ago, things would now be different - and better - than they actually are. But I cannot say this in the indicative mood: ‘If I invested in Western Mining ten years ago, things are different from the way they actually are’ is nonsense. Or consider the difference between ‘If Oswald had not shot Kennedy, Kennedy would have won a second term’ and ‘If Oswald did not shoot Kennedy, Kennedy won a second term’. The reason we reject the second, despite accepting the first, is that we know that the way things actually are Kennedy did not live to win a second term. Subjunctive conditionals typically concern non-actual worlds, whereas indicative conditionals concern the actual world under various hypotheses about what it is like. Thus we say that if Oswald did not shoot Kennedy, someone else did, because that is the only thing to think about the actual world (in which we know that Kennedy was shot) given that Oswald was not responsible.

3 Adams and Lewis

What about the acceptance or justified assertion conditions, as opposed to truth-conditions, for indicative conditionals? Many, including particularly Ernest Adams (1975), have urged that it is justified to assert ‘If A then C ’ to the extent that C is probable given A . Because the probability of C given A - in symbols, $Pr(C|A)$ - is (roughly) the probability we ascribe to C on learning A (see [Probability, interpretations of §5](#)), this suggestion fits nicely with the plausible idea that one is prepared to assert ‘If A then C ’ to the extent that one is prepared to infer C on learning A . It also explains our reluctance to assert together ‘If A then C ’ and ‘If A then not- C ’ when A is consistent, for when A is consistent, $Pr(C|A) = 1 - Pr(\text{not-}C|A)$, so they cannot be high together.

Can we explain this assertibility condition in terms of truth-conditions? David Lewis (1976) showed that it cannot be the case that the truth-conditions of ‘If A then C ’ are such that $Pr(\text{If } A \text{ then } C) = Pr(C|A)$. For then this equality would hold for all probability functions Pr , and that leads to trouble as follows:

$$Pr(\text{If } A \text{ then } C) = Pr(\text{If } A \text{ then } C|C).Pr(C) + Pr(\text{If } A \text{ then } C|\text{not-}C).Pr(\text{not-}C),$$

by expansion by cases. But if $Pr(\text{If } A \text{ then } C) = Pr(C|A)$ holds for all Pr , then it holds for $Pr(-|C)$ and $Pr(-|\text{not-}C)$, as the class of probability functions is closed under conditionalization. This means that

$$Pr(\text{If } A \text{ then } C|C) = Pr(C|A.C)$$

$$Pr(\text{If } A \text{ then } C|\text{not-}C) = Pr(C|A.\text{not-}C)$$

(where $Pr(C|A.C)$ is the probability of C given the conjunction of A and C). Hence we have

$$\begin{aligned} Pr(\text{If } A \text{ then } C) &= Pr(C|A.C).Pr(C) + Pr(C|A.\text{not-}C).Pr(\text{not-}C) \\ &= 1.Pr(C) + 0.Pr(\text{not-}C) \\ &= Pr(C) \end{aligned}$$

But then, by the claim under discussion, $Pr(C) = Pr(C|A)$. This is a ‘reductio’, for in general the probability of C is not independent of that of A . (This is the essence of the simplest of the proofs Lewis offered. For the substantial developments see Lewis (1986).)

Many respond to this proof by holding that indicative conditionals do not have truth-conditions at all; they only have assertion or acceptance conditions. This theory undoubtedly has its attractions but faces two problems. What

should be said about the powerful intuition that a conditional with a true antecedent and a false consequent is false? And how should the relevant notion of assertibility be elucidated? Not in terms of likelihood of truth, obviously.

An alternative strategy, called the supplemented equivalence theory, is to return to the equivalence thesis and argue that there is a convention governing the assertion of 'If A then C ' to the effect that it should only be asserted when it would be right to infer C on learning A . This convention is like that governing the use of 'but'. 'A but C ' has the same truth-conditions as 'A and C ', but the use of the former conventionally implicates, in H.P. Grice's terminology, a contrast (see Grice 1975; [Implicature](#)). Likewise, runs the suggestion, 'If A then C ' has the same truth-conditions as ' $A \supset C$ ' but its use carries the implicature that my reasons for ' $A \supset C$ ' are such that it would be right, on learning A , to infer C (that is, to use *modus ponens*). Now that will be the case just if the probability of ' $A \supset C$ ' would not be unduly diminished on learning that A is true - otherwise it would not then be available as a probably true premise to combine with A on the way to inferring C . It follows that it will be right to assert 'If A then C ' to the extent that (1) ' $A \supset C$ ' is probable, and (2) ' $A \supset C$ ' is probable given A . It is an elementary exercise in probability theory to show that this two-fold condition is satisfied to the extent that $Pr(C|A)$ is high. The supplemented equivalence theory therefore explains the assertibility condition noted by Adams.

What do supplemented equivalence theorists say about our earlier example 'If I live in London, then I live in Scotland', which comes out true on the equivalence theory and so on the supplemented equivalence theory? They say that it seems false not because it is false but because it has very low assertibility - $Pr(\text{I live in Scotland}|\text{I live in London}) = 0$ - and in general that our intuitions about truth and validity of inferences involving indicative conditionals are governed by responses to assertibility rather than truth.

Warning: almost everything about indicative conditionals is controversial, including whether they are well labelled by the term 'indicative', and some even deny the validity of *modus ponens*!

See also: [Logical and mathematical terms, glossary of](#); [Relevance logic and entailment](#)

FRANK JACKSON

References and further reading

- Adams, E.** (1975) *The Logic of Conditionals*, Dordrecht: Reidel. (A detailed defence of an approach to conditionals that denies them truth-conditions and shows how to develop a logic of conditionals based on his views about their assertibility.)
- Appiah, A.** (1985) *Assertion and Conditionals*, Cambridge: Cambridge University Press. (A detailed exposition of an account of indicative conditionals in terms of assertibility, set within the philosophy of mind.)
- Edgington, D.** (1995) 'On Conditionals', *Mind* 104: 235-329. (Excellent survey article that contains interesting material in its own right; includes a comprehensive and up-to-date reading list.)
- Grice, H.P.** (1975) 'Logic and Conversation', in F. Jackson (ed.) *Conditionals*, Oxford: Oxford University Press, 1991. (Explains implicature and the distinction between conversational and conventional implicature.)
- Jackson, F.** (1987) *Conditionals*, Oxford: Blackwell. (A detailed defence of the supplemented equivalence theory with criticisms of various alternative accounts of indicative conditionals.)
- Lewis, D.K.** (1973) *Counterfactuals*, Oxford: Blackwell. (The book which, along with Stalnaker's paper, launched the possible worlds approach to conditionals.)
- Lewis, D.K.** (1976) 'Probabilities of Conditionals and Conditional Probabilities', in F. Jackson (ed.) *Conditionals*, Oxford: Oxford University Press, 1991. (Gives the full version of the proof sketched in this entry.)
- Lewis, D.K.** (1986) 'Probabilities of Conditionals and Conditional Probabilities II', in F. Jackson (ed.) *Conditionals*, Oxford: Oxford University Press, 1991. (More on the proof sketched in this entry.)
- Stalnaker, R.** (1968) 'A Theory of Conditionals', in F. Jackson (ed.) *Conditionals*, Oxford: Oxford University Press, 1991. (The paper which, along with Lewis's book, launched the possible worlds approach to conditionals.)

Indirect discourse

Indirect discourse is a mode of speech-reporting whereby a speaker conveys the content of someone's utterance without quoting the actual words. Thus, if Pierre says, 'Paris est belle', an English speaker might truly say,

(1) Pierre said that Paris is beautiful.

In English, sentences of indirect discourse often have the form 'A said that s', where 'A' refers to a person and 's' is often called the 'content sentence' of the report.

Sentences of indirect discourse have been classed with attributions of belief (and other psychological states) in view of an apparent conflict with the 'principle of the intersubstitutability of coreferring terms', which states that the truth-value of a sentence does not alter if one term in a sentence is replaced with another referring to the same thing. If (1) is true and 'Paris' and 'the City of Light' refer to the same thing, (2) may still be false:

(2) Pierre said that the City of Light is beautiful.

1 Frege's theory

In 'indirect discourse' a speaker reports someone's utterance without quoting the actual words used (often using the form 'A said that s'). Sentences of indirect discourse give rise to an apparent conflict with the 'principle of the intersubstitutability of coreferring terms', which states that the truth-value of a sentence does not alter if one term is replaced with another which refers to the same thing. Gottlob Frege (1892) put forward an important theory to deal with this problem. He proposed that 'said' alters what the words in the 'content sentence' ('s' above) refer to. For example, consider the following statements:

(1) Pierre said that Paris is beautiful.

(2) Pierre said that the City of Light is beautiful.

'Paris' and 'the City of Light' both usually refer to the actual city, Paris. The phrase 'is beautiful' usually refers to the set of beautiful things. But when words appear after the phrase 'said that', they refer to the meanings they normally express - which Frege called their 'senses' - which are not the same as their referents (see Frege, G. §3).

The sense of an expression is a 'mode of presentation' of its referent. The sense associated with 'the City of Light' presents Paris in a particular way: as the City of Light. To understand an expression is to grasp its sense (see Sense and reference). Frege called the sense of a sentence a 'thought'. To have a propositional attitude is to stand in a certain relation to a thought. For example, to believe that Paris is beautiful is to stand in the believing relation to the thought that Paris is beautiful. Frege claimed that senses and thoughts are not psychological, mind-dependent things, but rather are Platonic, abstract objects. If two people believe that Paris is beautiful, then it is the very same thought that they both have. And this thought would exist even if there had been no minds (see Frege, G. §4).

The Fregean theory explains away the apparent failures of intersubstitution: the 'senses' of 'Paris' and 'the City of Light' are different. Hence the words refer to different things in (1) and (2) and so no violation of the substitution principle occurs.

2 Davidson's theory

Donald Davidson (1968/69) has provided a very different account of indirect discourse. Davidson proposed that, at the level of logical form (see Logical form), the analysis of speech reports should be composed of two separate sentences, the first ending in the demonstrative pronoun 'that'. The analysis of (1) above would then be:

(3) Pierre said that. Paris is beautiful.

On Davidson's theory the two sentences are logically distinct (not connected by any logical particle, such as 'and'). However, on specific occasions of use the demonstrative 'that' ending the first sentence refers to the reporter's utterance of the second.

According to this theory, the direct object of the verb 'said' is the reporter's own token utterance. So, suppose that

someone, *A*, utters (1). This speech act is composed of two token utterances: one of ‘Pierre said that’ and one of ‘Paris is beautiful’. Call this latter utterance ‘*U*’. Then, what Pierre would be reported as saying is *U*.

This feature is surprising, since it does not seem that Pierre could have said *U* (the reporter’s utterance) itself. But Davidson provides an account of what ‘said’ means which addresses the difficulty. He introduces the notion of ‘samesaying’: roughly, two individuals are samesayers if and only if they produce utterances which have relevantly similar semantic contents. Davidson has provided a general theory of meaning and interpretation which underwrites his appeal to semantic contents in this context (1984). We are then to understand (1) along the following lines:

(4) There was some utterance, *u*, of Pierre’s, and *u* and my next utterance make us samesayers. Paris is beautiful.

The idea is that the relation expressed by ‘said’ in *A*’s utterance of (1) can be understood in terms of the more complex relation spelled out in (4). But (4) is not intended to be any part of a technical semantic theory. It is just an informal, extra-theoretic guide to readers to help them understand the theory. In a semantic theory in Davidson’s style (see Davidson, D. §4), the axiom for ‘said’ might be:

(5) $(x)(y)(\text{‘said’ is true of } \langle x, y \rangle \text{ iff } x \text{ said } y)$.

The point is that since ‘said’ is supposed to relate reportees to utterances of reporters, some informal explanation of this saying relation is required. (4) provides this explanation.

Davidson’s theory explains the apparent failures of substitution such as that illustrated by (1) and (2). In an utterance of (1) the main verb is ‘said’ and its syntactic object is the demonstrative ‘that’. The ensuing utterance of ‘Paris is beautiful’ only serves as a referent for the demonstrative. It is not in any other way a semantically relevant component of the report. An utterance of (1) and an utterance of (2) resemble each other in that each has the form ‘Pierre said that’. But the referents of the two instances of ‘that’ are different utterances. So (1) does not entail (2). No violation of the intersubstitutability of coreferring terms occurs. The latter principle still applies to the content sentences of (1) and (2):

(6) Paris is beautiful.

(7) The City of Light is beautiful.

(6) does entail (7), given the coreference of ‘Paris’ and ‘the City of Light’. But this has no direct bearing on the logical relations between (1) and (2).

Davidson’s theory differs importantly from Frege’s. The former does not entail any shifts of reference for the words in the content sentence, which has exactly the same semantic features as it would have if uttered in isolation. This feature of the theory implements a natural and elegant idea of how speech reporting works. When *A* reports on the speech of *B*, *A* produces a new sentence, *S*, as content sentence of the report. *S* in *A*’s mouth expresses the content of *B*’s speech act. The utterance of *S* then serves as a model of *B*’s original speech act. And this exhausts its role.

Various objections have been made to Davidson’s proposal. Only two are discussed here. The first objection is due to Tyler Burge (1986). Burge points out that on Davidson’s theory, an utterance of (8) below would be true only if there were some utterance of the reporter’s and some utterance of Pierre’s that made them samesayers:

(8) Pierre said something.

This result follows from the functioning of the word ‘said’ on the theory. But the result is wrong: I might utter (8) truly without myself ever making any utterance that would make Pierre and me samesayers. The objection is perhaps not conclusive. But Burge points out that a solution to the difficulty would be to allow the objects of speech reports to be abstract objects, rather than token utterances, and offers several further arguments that this solution is correct. It seems natural to take the relevant abstract objects to be sentence-types. If this suggestion is adopted, then Burge’s objection forces no major revisions of Davidson’s theory. The original account of ‘said’ in terms of samesaying no longer applies. But an analogous explanation of the functioning of ‘said’ remains available. Instead of understanding (1) along the lines of (4), we would understand it along the lines of:

(9) There was some sentence, *s*, which Pierre uttered, and *s* meant in Pierre's mouth then what the following means now in mine: Paris is beautiful.

The second objection is due to James Higginbotham (1986). On Davidson's account, (10) and (11) ought to have identical semantic features:

(10) Every girl said that her mother was kind.

(11) Every girl said that. Her mother was kind.

This is because (11) merely makes explicit the analysis that Davidson would attribute to (10). However, (10) and (11) are semantically distinguishable. (10) is ambiguous: on the first reading, (10) would be true if and only if each girl said of some specific contextually indicated person, *x*, that *x*'s mother was kind; on the second reading, it would be true if and only if each girl said that her own mother was kind. But (11) is only subject to the first reading. This objection shows that Davidson's theory is not correct as it stands, and will have to be modified in very serious ways if it is to accommodate interactions between words inside and outside content sentences.

See also: [Logical and mathematical terms, glossary of](#); [Propositional attitude statements](#)

GABRIEL SEGAL

References and further reading

- Burge, T.** (1986) 'On Davidson's "On Saying That"', in E. Lepore (ed.) *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, New York: Blackwell, 191-208. (A critical article on Davidson's theory of indirect discourse.)
- Davidson, D.** (1968/69) 'On Saying That', *Synthese* 19: 130-46. (The classic work of Davidson's theory of indirect discourse, including criticisms of Frege's theory.)
- Davidson, D.** (1984) *Inquiries into Truth and Interpretation*, Oxford: Clarendon Press. (A collection of Davidson's articles in the philosophy of language, including some seminal works on the theory of meaning and interpretation.)
- Frege, G.** (1892) 'Über Sinn und Bedeutung', *Zeitschrift für Philosophie und philosophische Kritik* 100: 25-50; trans. M. Black, 'On Sense and Reference', in *Translations from the Philosophical Writings of Gottlob Frege*, ed. P.T. Geach and M. Black, Oxford: Blackwell, 1962, 56-78; 3rd edn, 1980. (The first clear statement of the problems of apparent failures of intersubstitutability from the standpoint of technical semantic theory. One of the most influential articles in philosophy of language.)
- Geach, P.T.** (1957) *Mental Acts*, London: Routledge & Kegan Paul. (A general work in the philosophies of mind and language, including extended discussion of the relationship between saying and propositional attitudes such as believing or judging.)
- Higginbotham, J.** (1986) 'Linguistic Theory and Davidson's Program in Semantics', in E. Lepore (ed.) *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, New York: Blackwell, 29-48. (Argues for the incorporation of Davidson's main technical innovations in philosophical semantic theory into empirical linguistics and includes a sophisticated theory of indirect speech that retains some core aspects of Davidson's proposal but avoids the main objections to it.)
- Scheffler, I.** (1954) 'An Inscriptional Approach to Indirect Quotation', *Analysis* 10: 83-90. (An influential nominalist account of indirect discourse.)
- Segal, G. and Speas, M.** (1986) 'On Saying *ŏČbepsi;t*', *Mind and Language* 1 (2): 124-32. (An article criticizing Davidson's theory of indirect discourse on empirical grounds.)
- Quine, W.V.** (1960) *Word and Object*, Cambridge, MA: MIT Press, 1964. (A general work in systematic philosophy including arguments critical of semantic theories, such as Frege's, that use a mentalistic or abstract notion of meaning; and discussion of various approaches to indirect discourse and propositional attitude statements.)

Inductive definitions and proofs

An inductive definition of a predicate R characterizes the R s as the smallest class which satisfies a basis clause of the form $(\beta(x) \rightarrow Rx)$, telling us that certain things satisfy R , together with one or more closure clauses of the form $(\Phi(x, R) \rightarrow Rx)$, which tell us that, if certain other things satisfy R , x satisfies R as well. 'Smallest' here means that the class of R s is included in every other class which satisfies the basis and closure clauses.

Inductive definitions are useful because of inductive proofs. To show that every R has property P , show that the class of R s that have P satisfies the basis and closure clauses.

The closure clauses tell us that if certain things satisfy R , x satisfies R as well. Thus satisfaction of the condition $\Phi(x, R)$ should be ensured by positive information to the effect that certain things satisfy R , and not also require negative information that certain things fail to satisfy R . In other words, the condition $\Phi(x, R)$ should be monotone, so that, if $R \subseteq S$ and $\Phi(x, R)$, then $\Phi(x, S)$; otherwise, we would have no assurance of the existence of a smallest class satisfying the basis and closure conditions.

While inductive definitions can take many forms, they have been studied most usefully in the special case in which the basis and closure clauses are formulated within the predicate calculus. Initiated by Yiannis Moschovakis, the study of such definitions has yielded an especially rich and elegant theory.

1 Inductive definitions in general

Inductive definitions occur throughout mathematics. For example, one aspect of the reduction of classical mathematics to set theory has been to treat natural numbers as sets, identifying a natural number with the set of its predecessors, so that they comprise the smallest class such that:

$\emptyset$ is a natural number

If $x = y \cup \{y\}$, for some natural number y , x is a natural number.

Having this inductive definition, we can give proofs by *mathematical induction*. To show that every number has property P , show that $\emptyset$ has the property and that, if y is a natural number with property P , so is $y \cup \{y\}$.

One inductive definition begets others. We define 'is a sum of' as the smallest ternary relation such that:

If x is a natural number, x is a sum of x and $\emptyset$.

If u is a sum of y and v , $x = u \cup \{u\}$ and $z = v \cup \{v\}$, then x is a sum of y and z .

An inductive proof shows that this stipulation uniquely determines a binary operation on the natural numbers.

We can extend arithmetical methods beyond the finite, identifying a limit ordinal with the set of its predecessors, so that the ordinals are the smallest class such that:

$\emptyset$ is an ordinal.

If $x = y \cup \{y\}$, for some ordinal y , then x is an ordinal.

If the set x is a union of ordinals, x is an ordinal.

Again one inductive definition leads to another. Given a function G , to define a function F on the ordinals such that:

$$F(\alpha) = G(\alpha, F \upharpoonright \alpha)$$

(where $F \upharpoonright \alpha = \{ \langle \beta, F(\beta) \rangle : \beta < \alpha \}$), take the smallest class of ordered pairs such that:

$\langle 0, G(0, \emptyset) \rangle$ is in F .

If h is a function defined on the ordinals $< \alpha$ such that $\langle \beta, h \upharpoonright \beta \rangle$ is in F , for each $\beta < \alpha$, then $G(\alpha, h)$ is in F .

Speaking abstractly, an inductive definition describes the smallest relation such that:

If $\beta(x_1, \dots, x_n)$, then $Rx_1 \dots x_n$.
 If $\Phi_1(x_1, \dots, x_n, R)$, then $Rx_1 \dots x_n$.
 ...
 If $\Phi_m(x_1, \dots, x_n, R)$, then $Rx_1 \dots x_n$.

Here β is an n -ary relation, and the conditions $\Phi_1, \dots, \Phi_m$ are monotone relations between n -tuples and n -ary relations.

Since the disjunction of two monotone conditions is a monotone condition, we can consolidate the basis and inductive clauses thus:

If $\beta(x_1, \dots, x_n)$ or $\Phi_1(x_1, \dots, x_n, R)$ or ... or $\Phi_m(x_1, \dots, x_n, R)$, then $Rx_1, \dots, x_n$.

Consequently, there will be no loss of generality if we only consider inductive definitions that describe the smallest class such that:

If $\Phi(x_1, \dots, x_n, R)$, then $Rx_1, \dots, x_n$,

where Φ is monotone.

Setting the example of the ordinals aside, let us restrict our attention to situations in which there is some set A such that, for each $x_1, \dots, x_n$ and S , if $\Phi_1(x_1, \dots, x_n, S)$, then each $x_i \in A$. Then, we shall not have to be concerned about the existence conditions for proper classes.

It is only because of the requirement that Φ is monotone that we can be confident that there is a smallest class which satisfies the closure clause, as we can see if we try to find the smallest class such that:

If either $x = 0$ and $\neg R1$ or $x = 1$ and $\neg R0$, then Rx .

Because of the monotony requirement, however, the intersection of all the sets which satisfy the closure clause satisfies the closure clause. It also satisfies the converse of the closure clause, so that (setting $n = 1$):

$(\forall x)(Rx \leftrightarrow \Phi(x, R))$.

We have inductive proofs: if we know $(\forall x)((Rx \wedge \Phi(x, P)) \rightarrow Px)$, we can conclude $(\forall x)(Rx \rightarrow Px)$.

Inductively defined classes are naturally arrayed in a well-ordered sequence of layers, for we can define the level of a member of R to be the number of applications of the closure clause that are needed to capture it. More precisely, we define, for each ordinal α , R_α to be $\{x: \Phi(x, \cup_{\beta < \alpha} R_\beta)\}$. Then the R_α s form a nonstrictly increasing sequence (that is, if $\beta < \alpha$, then $R_\beta \subseteq R_\alpha$); their union is R . Because we have stipulated that the R_α s are all contained within a set A , the sequence cannot be strictly increasing, since, if each R_α introduced new members of R , there would be as many members of R as there are ordinals. Consequently, there must exist an ordinal α with cardinality at most the cardinality of A such that $R_\alpha = R_{\alpha+1} = R$. The least such α is the *height* of the inductive definition.

2 First-order positive inductive definitions

An occurrence of a predicate R in a formula of the predicate calculus without ' $\rightarrow$ ' or ' $\leftrightarrow$ ' is *positive* if it occurs within the scope of an even number of negation signs. A first-order formula $\Phi(x_1, \dots, x_n, R)$ in which R occurs only positively describes a monotone condition. *First-order positive inductive definitions*, in which the antecedent of the closure clause is a first-order formula in which the defined term occurs only positively, have been extensively and fruitfully studied.

Let A be a model for a language $\mathcal{L}$ for the first-order predicate calculus, and let $\mathcal{L}_A$ be the language derived from $\mathcal{L}$ by adding a new constant $\underline{a}$ for each element a of the universe A of A . Apart from the predicate being defined, our first-order positive inductive definitions will consist of symbols of $\mathcal{L}_A$.

We can explicitly describe how an inductively defined set is built. Suppose C is inductively defined as the smallest set such that:

If $\Phi(x, R)$, then Rx .

Then a is in C if and only if $R\bar{a}$ is derivable from the axiom:

$$(\forall x)(\Phi(x, R) \rightarrow Rx)$$

in A -logic, that is to say, $R\bar{a}$ is derivable from the axiom together with atomic and negated atomic sentences true in A by first-order logic supplemented by the A -rule: from $\psi(\bar{a})$, for each a in A , to infer $(\forall x)\psi(x)$.

We shall restrict our attention to cases in which A is *acceptable*, that is, A contains an isomorphic copy of the natural number system and A can describe the formation of finite sequences. Acceptability of A will make it possible for $\mathcal{L}_A$ to describe its own syntax. We also suppose that $\mathcal{L}$ is built up from a finite vocabulary. Similar but more complicated analogues to the results described here hold without these restrictions.

It turns out that, to get a satisfactory theory, we need to look, not at the separate behaviour of individual inductive definitions but at their combinations. A relation C is *inductively definable* (or simply *inductive*) over A if C can be defined in a language derived from $\mathcal{L}_A$ by adding one or more inductively defined predicates; in the formula defining C the new predicates are allowed to occur only positively. More concisely, C is inductively definable over A if and only if there exist an $n + m$ -ary relation D and members $b_1, \dots, b_m$ of the universe A of A such that D is defined by a first-order positive inductive definition and C is equal to

$\langle a_1, \dots, a_n \rangle : \langle a_1, \dots, a_n, b_1, \dots, b_m \rangle \in D$. Showing this equivalence is a matter of seeing how several inductive definitions can be consolidated.

An inductively defined set whose complement is also inductively defined is *hyperclementary*.

A set C is inductive if and only if there is a finitely axiomatized theory Δ and formula $\psi(x)$ in a language extending $\mathcal{L}_A$ such that, for any a , a is an element of C if and only if $\psi(\bar{a})$ is derivable from Δ in A -logic. Just in case C is hyperclementary, we can arrange things so that, if a is in C then $\psi(\bar{a})$ is derivable from Δ , whereas if a is not in C , $\neg\psi(\bar{a})$ is derivable.

The results cited in the last paragraph would also obtain if the word ‘inductive’ were replaced by ‘recursively enumerable’, ‘hyperclementary’ by ‘recursive’ and ‘derivable in A -logic’ by ‘derivable in first-order logic’. This observation illustrates a far-reaching analogy between structural properties of inductively definable and recursively enumerable sets.

Another example is the following: inductive sets, like recursively enumerable sets, enjoy the *reduction property*, that is, for any inductive sets B and C , there exist nonoverlapping inductive sets $B' \subseteq B$ and $C' \subseteq C$ with $B' \cup C' = B \cup C$. We get them by comparing the stages in the relevant inductive definitions at which the members of B and C are introduced, putting a member of $B \cap C$ into B' just in case it is put into B before being placed into C .

Again, there are inductively defined sets C which are *complete*, meaning that, for each inductive set D , there is a hyperclementary function F such that, for any a , $a \in D$ if and only if $F(a) \in C$. The best-known complete set, at least among philosophers, is the minimal fixed point of Saul Kripke’s theory of truth. The theory of truth is robust enough to encode every inductive definition.

There is a version of the recursion theorem for inductive sets. Also, one can do priority arguments (see [Sacks 1990](#)).

The height of the monotone operator which generates an inductively defined set is a measure of the complexity of the set. Indeed, the principal ideas of the theory of inductive sets originate in S.C. Kleene’s investigations of how to extend the arithmetical hierarchy into the transfinite. The *order* of A , $o(A)$, is the supremum of the heights of all first-order positive inductive definitions over A . This supremum is reached by the complete inductive sets, among others. A subset of A is hyperclementary if and only if it is defined by a first-order positive inductive definition whose height is less than $o(A)$.

3 $\prod_1^1$ and inductively defined sets

An element a of A will be in the smallest class which satisfies the closure clause $(\Phi(x, R) \rightarrow Rx)$ just in case the sentence $(\forall R)((\forall x)(\Phi(x, R) \rightarrow Rx) \rightarrow R\bar{a})$ is true. It follows that every inductive set is $\prod_1^1$, that is, it is

definable by a formula consisting of initial second-order universal quantifiers followed by a first-order formula.

If A is countable, the converse holds. $(\forall R)\psi(R, \underline{a})$ is true if and only if the sentence $\psi(R, \underline{a})$ is a member of the smallest class of sentences which:

- contains the atomic and negated atomic sentences true in A ;
- contains the theorems of first-order logic;
- contains θ whenever it contains $(\chi \rightarrow \theta)$ and χ ; and
- contains $(\forall v)\chi(x)$ whenever it contains each $\chi(\underline{a})$.

Because A is acceptable, so that it can represent its own syntax, this can be written as a first-order positive inductive definition.

One proves this by proving a completeness theorem, showing that, for A countable, $(\forall R)\theta(R)$ is true if and only if $\theta(R)$ is derivable in A -logic.

4 Inductive and admissible sets

To get a picture of the place of inductively defined sets within the universe of set theory, we look at sets that can be built up by taking the elements of A as basic building blocks or *urelements*:

- $V_0 = A$.
- $V_{\alpha+1}$ = the set of subsets of V_α .
- $V_\lambda = \bigcup_{\alpha < \lambda} V_\alpha$, for λ a limit.

The *constructible sets* form a subclass of the V_α s:

- $L_0 = A$.
- $L_{\alpha+1}$ = the set of subsets of L_α which are definable in the language derived from $\mathcal{L}_A$ by adding a predicate ' U ', whose extension is A , a predicate ' $\in$ ', whose extension is the elementhood relation, restricted to L_α , and names for all the members of L_α .
- $L_\lambda = \bigcup_{\alpha < \lambda} L_\alpha$, for λ a limit.

We are interested in the initial stages of the constructible hierarchy. First, let us introduce a little more notation. The *bounded formulas* are those in which every universal quantifier takes the form $(\forall x)(x \in t \rightarrow \dots)$, and every existential quantifier takes the form $(\exists x)(x \in t \wedge \dots)$. The Σ_1 *formulas* are those obtained by prefixing one or more existential quantifiers to a bounded formula.

One can show that $L_{0(A)}$ constitutes the smallest set which:

- (1) contains A and the sets definable in A ;
- (2) contains x, y and the union of x whenever it contains x and y ;
- (3) satisfies the universal closure of the separation axiom schema, $(\exists y)(\forall z)(z \in y \leftrightarrow (z \in x \wedge \phi(z)))$, restricted to bounded formulas ϕ ; and
- (4) satisfies the universal closure of $(\forall x)(x \in t \rightarrow (\exists z)\phi(x, z)) \rightarrow (\exists y)(\forall x)(x \in t \rightarrow (\exists z)(z \in y \wedge \phi(x, z)))$, for ϕ bounded.

Subsets of the class of V_α s which satisfy these four conditions are said to be *admissible* above A . The first-order axioms which describe, in a natural way, a well-founded set which satisfies conditions (2) through (4) are called 'KPU' (for 'Kripke-Platek set theory with urelements', after the thinkers who first studied admissible sets).

The Barwise-Moschovakis-Gandy theorem states that a subset of A is inductive if and only if it is definable by a Σ_1 formula on $L_{0(A)}$. While the proof is complicated, it is possible to give a summary. To show that every inductively definable set is Σ_1 definable on $L_{0(A)}$, show that, if there is a proof in A -logic of a given sentence, then such a proof exists in $L_{0(A)}$. For the converse, show that, if S is Σ_1 definable, there is a formula σ such that $a \in S$ if and only if $\sigma(\underline{a})$ is derivable from KPU in A -logic.

It follows that a set is hyperelementary over A if and only if it is an element of $L_{0(A)}$.

The Barwise-Moschovakis-Gandy theorem explains the parallels between recursively enumerable and inductive sets. Recursively enumerable = Σ_1 definable in the smallest set admissible above $\emptyset$. Inductive over $A = \Sigma_1$ definable in the smallest set admissible above A .

Admissible sets have found widespread applications to the study of infinitary logic, but here it is possible to do no more than refer to Barwise (1975).

The results cited here are standard, and proofs can be found in Moschovakis (1974) and Barwise (1975).

See also: Constructible universe; Definition; Infinitary logics; Logical and mathematical terms, glossary of; Turing reducibility and Turing degrees

VANN McGEE

References and further reading

- Barwise, J.** (1975) *Admissible Sets and Structures: An Approach to Definability Theory*, Berlin and New York: Springer.(A clear and readable development of the theory of admissible sets, with particular attention to infinitary logic.)
- Moschovakis, Y.N.** (1974) *Elementary Induction on Abstract Structures*, Amsterdam: North Holland.(This is the fundamental work on inductively defined sets. Lucid and readable.)
- Sacks, G.E.** (1990) *Higher Recursion Theory*, New York: Springer-Verlag.(Uses the methods of recursion theory, particularly priority arguments, to describe the structure of inductively defined sets.)

Inference, Indian theories of

The use of argument in rational inquiry in India reaches almost as far back in time as its oldest extant literature. Even in very early texts, one finds the deliberate use of modus tollens, for example, to refute positions thought to be false. In light of such practice, it is not surprising to discover that Indian thinkers came to identify certain forms of reasoning and to study them systematically.

The study of inference in India is, as Karl Potter (1977) has emphasized, not the study of valid reasoning as reflected in linguistic or paralinguistic forms, but the study of the circumstances in which knowledge of some facts permits knowledge of another fact, and of when acceptance by one person of some state of affairs as a fact requires that that person accept another as a fact. Still, the form of inference which came to be systematically investigated in India can be given schematically (see below).

At the core of the study of inference in India is the use of a naïve realist ontology. The world consists of individual substances or things (dravya), universals (sāmānya), and relations between them. The fundamental relation is the one of occurrence (vṛtti). The relata of this relation are known as substratum (dharmin) and superstratum (dharma) respectively. The relation has two forms: contact (saṃyoga) and inherence (samavāya). So, for example, one individual substance, say a pot, may occur on another, say the ground, by the relation of contact. In this case, the pot is the superstratum and the ground is the substratum. Or a universal, say brownness, may occur in an individual substance, say a pot, by the relation of inherence. Here, brownness, the superstratum, inheres in the pot, the substratum. The converse of the relation of occurrence is the relation of possession.

Another important relation is the relation that one superstratum bears to another. This relation, known as pervasion (vyāpti), can be defined in terms of the occurrence relation. One superstratum pervades another just in case wherever the second occurs the first occurs. The converse of the pervasion relation is the concomitance relation. As a result of these relations, the world embodies a structure: if one superstratum H is concomitant with another superstratum S, and if a particular substratum p possesses the former superstratum, then it possesses the second. This structure is captured in this inferential schema:

Pakṣa (thesis): p has S.

Hetu (ground): p has H.

Vyāpti (pervasion): Whatever has H has S.

Here are two paradigmatic cases of such an inference:

Pakṣa (thesis): p has fire.

Hetu (ground): p has smoke.

Vyāpti (pervasion): Whatever has smoke has fire.

Pakṣa (thesis): p is a tree (that is, has tree-ness).

Hetu (ground): p is an oak (that is, has oak-ness).

Vyāpti (pervasion): Whatever is an oak (that is, has oak-ness) is a tree (that is, has tree-ness).

1 Inference and rational inquiry

In tracing out the origin and development of the science of reasoning in general, and of inference (*anumāna*) in particular, one must be careful to distinguish the exercise and implicit recognition of forms of reasoning in thought and inquiry from their explicit articulation and systematic study. This distinction is well exemplified by classical Greek mathematicians: the forms of arguments used in their proofs far exceeded Aristotle's syllogistic and Stoic propositional logic combined. In a similar way, classical Indian thinkers did not articulate many forms of argument which they instinctively recognized as valid and routinely deployed in rational inquiry.

While such principles of logic as the law of non-contradiction, the law of double negation, the law of excluded middle, De Morgan's laws and the equivalence of a proposition and its contrapositive were widely and routinely used by Indian thinkers, their formulation was not taken up for systematic study. Hence, the use of these principles of logic is not discussed in the present entry. Instead, attention is confined to how Indian thinkers came to identify

inference as captured in the inferential scheme and how they made such inferences the object of systematic study.

The history of thought in India can be divided into three periods: the pre-classical period (up to the fourth century BC), the classical period (up to the tenth century AD) and the scholastic period (from the eleventh century on). During the pre-classical period the practice of rational inquiry and debate developed, furnishing the raw material for the formulation of works on the science of reason - *hetu-śāstra* or *nyāya*, as the Indian thinkers called it - which appeared at the beginning of the classical period and played an important role in every major school of philosophical and empirical thought.

Rational inquiry comprises the search for reasons for publicly accepted facts, subject to public and rational scrutiny. It is an activity which involves people both severally and collectively. It involves people severally in so far as individual people are the locus of the exercise of reason. It involves people collectively in so far as the exercise of reason is sharpened by the scrutiny of others. Inference is central to both its individual and collective exercise.

Undoubtedly, the Indian science of reason has its origins in the emergence of rational inquiry and debate. However, the origins of rational inquiry and debate are not clear, the relevant literature being scant and the references therein scanty. None the less, a sketch of the earliest period can be constructed.

2 Seeds of rational inquiry

The seeds of rational inquiry in India are found in appendices to its earliest extant body of literature, the Vedas (Wisdom, Knowledge), composed between the thirteenth and eleventh centuries BC, and comprising four collections (*samhitās*): the *Rg Veda* (*Wisdom of the Verses*), hymns dedicated to the gods worshipped by the Indo-Aryans; the *Yajur Veda* (*Wisdom of the Sacrificial Formulas*), sacred formulas recited at rituals; the *Sāma Veda* (*Wisdom of the Chants*), an anthology of chants used by priests in their rituals; and the *Atharva Veda* (*Wisdom of the Atharva Priests*), a compendium of hymns, incantations and magic charms used by priests.

Their appendices include the *Brāhmaṇas* (Books Pertaining to Prayer, dating from between the eleventh and eighth centuries BC), and the *Āraṇyakas* (Books Studied in the Forest, from between the ninth and seventh centuries BC). In them one finds speculation about the causal relations involved in sacrificial rites. Thus, one encounters the concept of *ṛta*, the natural order of things, which underwrites the efficacy of sacrifice, that is, the capacity of the performance of sacrificial rites to bring about prescribed effects. One also encounters the concept of karma, whereby deeds performed at one point in one's life have consequences which manifest themselves in terms of pain and pleasure experienced later in one's life or in one's later life (see [Karma and rebirth, Indian conceptions of](#)). Moreover, speculation about the origins of the universe is also found in the *Aitareya Āraṇyaka*.

The kind of speculation just described appears to have become the object of public contests in which contestants showed off their knowledge of ritual and their proto-cosmological speculations as to its basis. Inevitably, the views of these contestants differed, and indeed conflicted. Such differences undoubtedly gave rise to questions of how to establish the correctness of one view and how to adjudicate between conflicting views. In this way, according to some scholars (for example, Witzel 1987), arose the practice of public debates, so common in the classical and scholastic periods.

In any event, over the course of time the range of objects of speculation grew to include not only questions of cosmology but also questions of metaphysics, especially the question of what there is, and the naturally ensuing question of how one knows or can justify one's claims about what there is. Thus, one finds views purporting to assist humankind in the alleviation of the unhappiness it suffers, views about what comprises happiness, what constitutes matter, whether or not there is an eternal soul, whether or not there is free will, and whether or not knowledge is possible. The diversity of views includes not only views defended in the Brahmanical literature, in particular the [Upaniṣads](#) (dating from the eighth to sixth centuries BC) and, in the Buddhist canon, the *Tipiṭika* (*Three Baskets*, dating from after the sixth century BC), but also dissenting views reported in the same literature, such as those of the materialist and atomist Ajita Kesakambalin, of the determinist Makkhali Gosāla, and of the sceptic Saṅgha Belatthiputta.

Central to the pursuit of these questions is the concept of causation, which itself fell within the purview of rational inquiry. Thus, for example, in the [Śvetāśvatara Upaniṣad](#) (1.2), a number of views pertaining to causation are

reported: determinism (*niyatīvāda*), indeterminism (*yadr̥cchāvāda*) and latency (*svabhāva*).

While the range of rational inquiry increased, it also deepened, for rational inquiry (*anvīkṣikī*, *parīkṣā*) itself became the object of rational inquiry, that is, questions were raised about how facts are known. Thus, one finds as early as the Taittirīya Āraṇyaka enumerations of the sources of knowledge (*pramāṇas*), namely perception (*pratyakṣa*), inference (*anumāna*), religious authority (*smṛti*) and tradition (*aitihya*).

In short, during the pre-classical period, three practices important to the development of the study of inference make their appearance. Rational inquiry into the causes of things, rational inquiry into how things are known, and public contests which perhaps anticipate public debates.

3 Growth of rational inquiry

By the fifth century BC, great social change was taking place in India and a period of intense intellectual activity began. Rational inquiry had taken hold and its topics were not confined to cosmology and soteriology. True, apart from Pāṇini's *Aṣṭādhyāyī*, the world's earliest extant grammar, no works devoted to such topics as agriculture, architecture, astronomy, grammar, law, mathematics, medicine, phonology and statecraft date from this period. None the less, scholars agree that this period contains the roots of many works not redacted until around the turn of the millennium, including *Kṛṣiśāstra* (*Treatise on Agriculture*), *Śilpāśāstra* (*Treatise on Architecture*), *Jyotiṣaśāstra* (*Treatise on Astronomy*), *Dharmaśāstra* (*Treatise on Law*), *Carakasamhitā* (*Caraka's Collection*, a medical treatise) and *Arthaśāstra* (*Treatise on Wealth*, a political treatise).

Allusions to questions about the number of sources of knowledge, their nature and their objects are found here and there in the **Upaniṣads** (for example, **Maitrī Upaniṣad**). Similar issues find elaboration in the Buddhist literature dating from the same period.

Many of the dialogues of the **Buddha**, Siddhārtha Gautama, show that some version of the law of non-contradiction is to be respected, that is, they show adherence to the view that consistency is a necessary condition for truth. Moreover, many dialogues suggest that truth consists in correspondence to facts, independent of observers. Indeed, a persistent message of the Buddha is that knowledge is grounded in personal observation and inference, where observation is taken to include paranormal observation (though observation none the less). In addition, Buddhists of this period recognize that inference itself must be grounded in causal explanation.

Indeed, much energy is devoted to the development of causal explanations, not only with respect to grounding soteriological claims, for example, as reflected in the doctrine of dependent origination (*pratītya-samutpada*), but even more broadly. Thus, for example, the *Paṭṭhāna* (Causes), one of the seven works of the *Abhidhamma piṭaka* (Basket of Further Doctrine), itself one of the three 'baskets' making up the Buddhist canon, classifies intentional states in terms of twenty-four kinds of causal relations. Causal relations are invoked to explain perceptual states as well. Indeed causal relations are routinely presupposed in discussions of physical events and processes, such as how seeds give rise to shoots (see **Causation, Indian theories of**).

Moreover, these thinkers not only used causal relations to construct explanations, they were clearly sensitive to the problem of how to establish the existence of causal relations. Passages in the Pāli Canon indicate awareness of the problem of drawing general conclusions from limited observations. Indeed, this period seems to have witnessed the formulation of Mill's method of agreement and difference, known to Indian thinkers as the method of *anvaya-vyatireka*. Such a method, as G. Cardona (1967) has shown, was clearly known to Kātyāyana (fl. third century BC), the author of the earliest extant commentary on the *Aṣṭādhyāyī*, and seems to have been known to the Lokāyatikas, who also date from this early period (see **Materialism, Indian school of**).

Publicly organized debates (*pariṣad*) were commonplace, alluded to in various **Upaniṣads** (**Chāndogya Upaniṣad 5.3.1**; **Bṛhadāraṇyaka Upaniṣad 6.2.1**) and exemplified in over half of the thirty-four *suttas* of the Buddhist *Dīgha Nikāya* (Collection of Long Discourses), which are a part of the *Sutta Piṭaka* (*Basket of Discourses*), one of the three 'baskets' of the Buddhist canon. A better-known, but much later, example of such engagements is the Buddhist work *Milindapañho* (*Questions of King Milinda*).

The earliest example of protracted systematic rational inquiry into metaphysical issues is found in Moggallīputta Tissa's *Kathāvatthu* (*Points of Controversy*), a Buddhist work of the third century BC. It sets out the refutation of some two hundred propositions over which the Theravādins, one of the Buddhist schools, disagreed with other

Buddhist schools. The treatment of each point comprises a debate between a proponent and an opponent. The refutations turn on demonstrating the inconsistency of a set of propositions.

Not only is the form of the debate canonical but the form of the reasoning is as well. Thus, for example, one finds the following pattern throughout Book 1, Chapter 1:

Proponent: Is A B ?

Opponent: Yes.

Proponent: Is C D ?

Opponent: No.

Proponent: Acknowledge defeat, since if A is B , then C is D .

It is clear that refutation hinges on inconsistency as exhibited by the following propositional schema: $p, \neg q, p \rightarrow q$. Though, as I. Bochenski (1956) has remarked, no propositional forms are identified as such, it is none the less clear that inconsistency of a certain form was considered a decisive refutation.

4 Appearance of the canonical inference

In the classical period, as stated above, begins the redaction of works codifying the results of systematic, rational inquiry begun earlier, not only into such fields as grammar, architecture, mathematics, law and politics, but also into the enduring philosophical questions, including the nature of rational inquiry itself. This is made clear in a brief passage in the *Arthaśāstra* (1.2), a work on statecraft attributed to Kauṭilya (fl. fourth century BC).

The texts surviving from this period have diverse origins. There are two important passages in the *Carakasamhitā*, a medical text which was redacted in its current form in the first century AD. There are also texts from the philosophical literature, both Brahmanical and non-Brahmanical. The best-known texts pertaining to inference are in a work attributed to Akṣapāda Gautama, the *Nyāyasūtra*, a Brahmanical treatise on rational inquiry, whose actual redaction is thought to date to the third century AD (see Gautama, Akṣapāda). Another Brahmanical work touching upon inference is the *Vaiśeṣikasūtra*, a work on speculative ontology attributed to Kaṇāda; this was redacted in its current form perhaps at the beginning of the second century AD, though perhaps some of its passages date back to the second century BC. Still another Brahmanical work, which survives only in fragments (which have been collected by E. Frauwallner), is a Sāṅkhya one entitled *Ṣaṣṭitantra* (Sixty Doctrines), attributed by some to Pāñcaśikha (fl. second century BC) and by others to Varṣanya (fl. after second century AD).

The remaining texts are found in the Buddhist philosophical literature. To begin with, there are four passages in four works by the Buddhist idealist Asaṅga (fl. fourth century AD): one in his *Abhidharmasangītiśāstra* (*Treatise on the Proclamation of the Abhidharma*), one in his *Abhidharmasamuccaya* (*Compendium of the Abhidharma*), one at the end of a chapter of his *Yogācārabhūmiśāstra* (*Treatise on the Stages of Yogic Practice*) and one in a work which survives only in Chinese translation, *Shun Zheng Lun* (*Treatise on According with What is Correct*). In addition, Vasubandhu (fl. fifth century AD), another Buddhist idealist, thought to be the younger brother of Asaṅga, wrote at least three works on debate: *Ru Shi Lun* (*Treatise on Resembling the Facts*, reconstructed in Sanskrit as *Tarkaśāstra*, or *Treatise on Reasoning*), which survives only in Chinese translation, the *Vādaavidhi* (*Rules of Debate*), whose Sanskrit fragments have been collected by E. Frauwallner, and the *Vādaavidhāna* (*Precepts of Debate*). Finally, there are two other works of unknown author and unknown date which survive only in Chinese translation: *Fang Bian Xin Lun* (reconstructed in Sanskrit by G. Tucci as *Upāyahṛdaya*, or *Essential Methods*), and *Xian Zhang Sheng Jiao Lun* (reconstructed in Sanskrit by G. Tucci as *Prakaraṇāryavācāśāstra*, or *Treatise on the Noble Words of Explanation*).

With the notable exception of the *Vaiśeṣikasūtra*, which treats inference only as an epistemic process, the greater part of each text is devoted to inference as argument in debate. These texts typically enumerate, define or classify public discussions, propositions as they are used in public discussions, parts of arguments, qualities which either enhance or detract from a discussant's performance, and statements or actions by a discussant which warrant his being considered defeated, including the uttering of various fallacies.

These texts identify the form of argument used in debate. Thus, for example, the *Nyāyasūtra* (1.1.32), like the *Carakasamhitā* (2.8.31), defines an argument as having five parts: the thesis (pakṣa), or proposition (*pratijñā*); the ground, or reason (*hetu*); the corroboration (*drṣṭānta*); the application (*upanaya*); and the conclusion (*nigamana*).

The *Carakasamhitā* furnishes the following example:

Thesis: The soul is non-eternal.

Ground: Because it is detectable by the senses.

Corroboration: It is like a pot.

Application: As a pot is detectable by the senses, and is non-eternal, so is the soul detectable by the senses.

Conclusion: Therefore, the soul is non-eternal.

This form of argument clearly reflects the debate situation. First, one propounds a proposition, that is, one sets forth a proposition to be proved. One then states the ground or reason for the proposition. Next, one corroborates with an example the implicit connection between the property mentioned in the proposition and the property adduced as its ground. The immediately ensuing step, the application, spells out the analogy between the example and the subject of the proposition. Notice that this part of the argument retains vestiges of the analogical reasoning that is no doubt its predecessor. Finally, one asserts the proposition.

As was obvious to these thinkers, not all arguments of this form are good arguments. Not surprisingly, the texts catalogue bad arguments. Grounds adduced in arguments catalogued as bad are referred to as non-grounds (*ahetu*) or pseudo-grounds (*hetu-ābhāsa*). It is difficult to be sure what the basis of the classification was. In the case of the *Nyāyasūtra*, the author gives neither a definition nor an example. Even in cases where definitions and examples are given, as in the *Carakasamhitā*, the modern reader is not always sure what is intended. In all likelihood, included here are cases where the premises of the argument can be true but the conclusion false - formal fallacies - and cases where the argument, though formally valid, is none the less unpersuasive, for example, because its ground (*hetu*) is as controversial as its conclusion.

These very same texts, as well as the *Vaiśeṣikasūtra*, touch on inference as an epistemic process. While the examples of inference furnished all comprise expressions corresponding with the thesis (*pakṣa*) and the ground (*hetu*) of the inferential schema, not all the texts are equally explicit in identifying the form of inference. In particular, both the *Carakasamhitā* (1.11.21-2) and the *Nyāyasūtra* (1.1.5) define inference as knowledge of one fact on the basis of knowledge of another, leaving unmentioned any knowledge of a relation linking the two. Moreover, these texts classify inferences on the basis of characteristics completely extrinsic to the logical features of the inferences adduced. Inferences appear to be classified according to the temporal order of the occurrences of the properties corresponding with those in the thesis (*pakṣa*) and the ground (*hetu*) of the inferential schema.

5 Formalization of the canonical inference

Improved definitions, which mention not only what corresponds with the thesis (*pakṣa*) and the ground (*hetu*) of the inferential schema, but also what corresponds with its pervasion (*vyāpti*), are found in the *Śaṣṭitantra*, the *Vaiśeṣikasūtra* and the *Abhidharmasamuccaya* (Further Doctrine Collection), where knowledge of a relation is explicitly included in their definitions of inference. However, what corresponds with pervasion (*vyāpti*) in the inferential schema is not a formal relation, but one of a miscellany of material relations instead. The *Śaṣṭitantra* enumerates seven such relations, while the *Vaiśeṣikasūtra* (9.20) enumerates five: the relation of cause to effect, of effect to cause, of contact, of exclusion, and of inherence. In each of these texts, the miscellany of material relations serves to classify inferences. Thus, although in these three works the parts of an inference are made explicit, the formal connection between them remains implicit.

As Katsura (1986) has observed, the earliest extant texts to grapple explicitly with making the formal connection between the parts of an inference apparent are those of Vasubandhu. To begin with, in his *Vādaśāstra*, Vasubandhu makes it clear that the relation, knowledge of which is necessary for inference, is not just any in a miscellany of material relations, but a formal relation, which he designates in some places as *avinābhāva* ('being a *sine qua non*') and in others as *nāntarīyakatva* ('being unmediated').

At the same time, Vasubandhu exploits the idea, ascribed by Asaōga in his *Shun Zheng Lun* to an unknown school (thought by at least one scholar to be the Sāṅkhya school), that a ground (*hetu*) in an inference is a proper one if and only if it satisfies three conditions - the so-called *trirūpahetu*, or the grounding property (*hetu*) in its three forms. The first form is that the grounding property should occur in the subject of an inference (*pakṣa*). The second is that the grounding property should occur in those substrata similar to the subject in so far as they have the

property to be established (*sādhya*). And third, the grounding property should not occur in any of those substrata dissimilar from the subject in so far as they lack the property to be established. The first form corresponds to the inferential schema's ground (*hetu*), while the second and the third correspond to its pervasion (*vyāpti*). With the formulation of the *trirūpahetu*, the formal nature of the classical Indian inference is at last fully circumscribed. This form is succinctly displayed by the following schema:

Pakṣa (thesis): *p* has *S*.

Hetu (ground): *p* has *H*.

Vyāpti (pervasion): Whatever has *H* has *S*.

An insightful distinction found in Vasubandhu's work is the one between inference for oneself and inference for others, which makes explicit what had previously been only implicit, namely that inference, the cognitive process whereby one increases one's knowledge, and argument, the device of persuasion, are but two sides of a single coin.

Dignāga (fl. beginning of the sixth century), another Buddhist philosopher, consolidates and systematizes the insights into the formal basis of inference found in Vasubandhu's works. First, he makes the distinction between inference for oneself and inference for others the organizational cornerstone of his treatment of inference. Second, he undertakes to make the three forms of the grounding property, the so-called *trirūpahetu*, more precise, pressing into service the Sanskrit particle *eva* (only). And third, and perhaps most strikingly, he devises the *hetucakra* ('wheel of reasons'), a three-by-three matrix set up to classify pseudo-grounds in the light of the last two forms of the *trirūpahetu*. On the one hand, there are the three cases of the grounding property occurring in some, none or all of the substrata where the property to be established occurs. On the other hand, there are the three cases of the grounding property occurring in some, none or all of the substrata where the property to be established does not occur.

6 Developments after formalization

Dignāga's works set the framework within which subsequent Buddhist thinkers address philosophical issues pertaining to inference and debate. Thus, Śāṅkarasvāmin (fl. sixth century) writes a brief manual of inference for Buddhists, called the *Nyāyapraveśa (Beginning Logic)*, based directly on Dignāga's work. Not long thereafter, **Dharmakīrti** (c.600-60), the great Buddhist metaphysician, also elaborates his views on inference and debate within the framework found in Dignāga.

Dharmakīrti makes at least two contributions to the treatment of inference. Recall that one of the developments found in Vasubandhu's work was the identification of the formal contribution of what corresponds with the inferential schema's pervasion (*vyāpti*), making it explicit that the corresponding relation is a formal one. One of Vasubandhu's terms for it, namely *avinābhāva*, or 'being a *sine qua non*', made it clear that inference involves some form of necessity. The question raised and systematically investigated by Dharmakīrti is: in what does this necessity consist? His answer is that it consists in two things: the causation relation (*tadutpatti*) and the identity relation (*tādātmya*). The second contribution by Dharmakīrti is his attempt to bring knowledge of absences, or roughly negative facts, within the purview of inference (see **Negative facts in classical Indian philosophy §§1-2**).

Another important Buddhist thinker who treated inference is Dharmottara (fl. eighth century). He wrote a useful commentary on Dharmakīrti's widely read *Nyāyabindu (Drop of Logic)*.

Dignāga not only had a profound influence on his Buddhist followers, he also influenced his non-Buddhist contemporaries and their followers. It would be wrong, however, to conclude that every adoption of ideas similar to those used by Dignāga in his works should be attributed to him. After all, we cannot be certain that his contemporaries did not arrive at similar ideas independently or get their ideas from sources common to them and Dignāga. Thus, for example, in his *Nyāyāvatāra (Descent into Logic)*, the Jaina thinker Siddhasena Divākara (c. fifth century), a contemporary of Vasubandhu, also speaks of the formal relation *antarvyāpti*, or inner pervasion, which corresponds to *vyāpti*, or pervasion, in the inferential schema. In addition, Praśastapāda (fl. sixth century), an adherent of the Vaiśeṣika school and near contemporary of Dignāga, also defines inference in a way which not only makes clear its formal nature but also uses the quantificational adjective *sarva* (all) to make the formal connection precise.

At the same time, some authors of this period seem to have retained a view of inference akin to the one found in the *Śaṣṭitantra* and the *Vaiśeṣikasūtra*, in which the formal role of what corresponds with the inferential schema's pervasion (*vyāpti*) has yet to have been identified. This is true both of **Vātsyāyana** (fl. fifth century), the author of the earliest extant commentary on the *Nyāyasūtra*, and of Śābara (fl. sixth century), the author of the earliest extant commentary on Jaimini's *Mīmāṃsāsūtra*.

However, it was not long before the advocates of both Nyāya and Mīmāṃsā adapted to the formal view of inference. On the one hand, one finds that the Mīmāṃsā thinker Kumārila Bhaṭṭa (fl. early seventh century) adopts, without special comment, the formal perspective. On the other hand, though the Nyāya thinker **Uddyotakara** (fl. late sixth century) argues vigorously against many of Dignāga's views, he none the less advocates a view which presupposes the formalization found in them. For example, Uddyotakara rejects the suitability of the quantifying particle *eva* (only) in the formulation of the three forms of the grounding property (*trirūpahetu*), yet he himself introduces a classification of inference which presupposes Dignāga's formalization. Thus, Uddyotakara classifies grounds (*hetu*) as: concomitant (*anvaya*), where nothing distinct from a particular substratum *p* (in the inferential schema) fails to have the property *S*; exclusive (*vyatireka*), where nothing distinct from *p* has *S*; and both concomitant and exclusive, where some things distinct from *p* have *S* and some fail to have *S*. This classification becomes the standard classification for the adherents of Nyāya during the scholastic period.

The developments of the scholastic period are confined almost entirely to the so-called Navya-Nyāya school. Its contributors to the treatment of inference include **Gaṅgeśa** (fl. fourteenth century), perhaps the most outstanding of all the thinkers of this school, Raghunātha (fl. early sixteenth century), Jagadīśa (fl. later seventeenth century) and **Gadādhara** (1604-1709).

We conclude with two warnings. First, the treatment of Indian inference here has centred exclusively on tracing out the unfolding of the inferential schema in the history of Indian thought. It has only hinted at the rich philosophical debates that arose among those Indian thinkers eager to harmonize their insights into the formal nature of inference with their views pertaining to metaphysical and linguistic questions. Second, the treatment has not prepared the reader for the intricacies and pitfalls of the terminology used by Indian thinkers in their discussions of inference. The terminology is both ambiguous and inconstant, not just as one goes from the work of one thinker to that of another, but even within the work of a single thinker. Annoying as the contemporary reader may find such ambiguity, it is necessary to remember that this is an inevitable by-product of intellectual fermentation.

See also: **Definition, Indian concepts of**; **Knowledge, Indian views of**; **Nyāya-Vaiśeṣika**

BRENDAN S. GILLON

References and further reading

- Annambhaṭṭa** (c.1600) *Tarkasamgraha*, trans. M.R. Bodas, Poona: Bhandarkar Oriental Research Institute, 1918. (A very clear manual of logic for the Nyāya school.)
- Bochenski, I.M.** (1956) *Formale Logik*, trans. I. Thomas, *A History of Formal Logic*, Notre Dame, IN: University of Notre Dame Press, 1961. (Part 6 is a forty-page history of Indian inference.)
- Cardona, G.** (1967) 'Anvaya and Vyatireka in Indian Grammar', *Adyar Library Bulletin* 5 (31-2): 313-52. (The author describes the origins of Mill's method of agreement and difference in early Indian thought.)
- Dharmakīrti** (c.650) *Nyāyabindu (Drop of Logic)*, trans. M. Gangopadhyaya, Calcutta: Indian Studies Past and Present, 1971. (A good introduction to Dharmakīrti's metaphysics and logic.)
- Dignāga** (c.525) *Nyāyamukha (Introduction to Systematic Reasoning)*, trans. G. Tucci, *The Nyāyamukha of Dignāga*, Heidelberg: Materialien zur Kunde des Buddhismus, 1930. (This very brief text summarizes Dignāga's views on inference and debate.)
- Gillon, B.S.** (1986) 'Dharmakīrti and his Theory of Inference', in B.K. Matilal and D. Evans (eds) *Buddhist Logic and Epistemology*, Dordrecht: Reidel, 77-87. (A self-contained explanation of how Dharmakīrti treats knowledge of absences as a form of inference.)
- Gillon, B.S. and Hayes, R.** (1982) 'The Role of the Particle *Eva* in (Logical) Quantification in Sanskrit', *Wiener Zeitschrift für die Kunde Südasiens* 26: 195-203. (The authors show how the Sanskrit particle *eva* ('only') is used to express logical quantification by Indian logicians.)
- Gillon, B.S. and Love, M.L.** (1980) 'Indian Logic Revisited: *Nyāyapraveśa* Reviewed', *Journal of Indian Philosophy* 8: 349-84. (The authors set out the logical and grammatical details of Indian inference as found in

the *Nyāyapraveśa*.)

- Hayes, R.** (1980) 'Dīṇāga's Views on Reasoning (*Svārthānumāna*)', *Journal of Indian Philosophy* 8: 219-77. (A thorough and scholarly treatment of Dignāga's theory of inference.)
- Jayatillike, K.N.** (1963) *Early Buddhist Theory of Knowledge*, London: Allen & Unwin. (Thorough survey of the intellectual climate at the time of the Buddha.)
- Katsura Shoryu** (1986) 'On the Origin and Development of the Concept of *Vyāpti* in Indian Logic', *Hiroshima Tetsugakkai* 38: 1-16. (The author traces the origin of the Indian form of inference.)
- Kauṭilya** (4th century BC) *Arthaśāstra*, ed. and trans. R.P. Kangle, Delhi: Motilal Banarsidass, 1988, 3 vols. (A work on statecraft.)
- Matilal, B.K.** (1985) *Logic, Language, and Reality: An Introduction to Indian Philosophical Studies*, Delhi: Motilal Banarsidass. (The first two chapters address many of the outstanding philosophical and philological problems pertaining to the study of inference in Indian philosophy. This is a difficult but important work by one of the twentieth century's major contributors to Indian philosophy.)
- Matilal, B.K. and Evans, D.** (eds) (1986) *Buddhist Logic and Epistemology*, Dordrecht: Reidel. (A collection of articles by experts on Buddhist logic and epistemology.)
- Milindapañho* (*The Questions of King Milinda*) (c.100 AD), The Sacred Books of the East 35-6, trans. T.W. Rhys-Davids, Oxford: Clarendon Press, 1890-4, 2 vols; repr. New York: Dover, 1963. (A work on Buddhist metaphysics set out as a dialogue between the senior Buddhist monk, Nāgasena, and the Indo-Greek king Menander.)
- Moggaliputta Tissa** (c.255 BC) *Kathāvatthu* (*Points of Controversy*), trans. S.Z. Aung and C.A.F. Rhys Davids, London: The Pali Text Society, 1915. (The author sets out the refutation of some 200 propositions over which the Sthavīravādins, one of the Buddhist schools, disagreed with the Sarvāstivādins, another Buddhist school. The text illustrates both the canonical forms of debate and reasoning.)
- Mokṣākaragupta** (c.1000) *Tarkabhāṣā*, trans. G. Jhā, Poona: Oriental Book Agency, 1967. (A very clear manual of logic for Buddhists.)
- Nyāyasūtra* (c.3rd century AD), trans. G. Jhā, Delhi: Motilal Banarsidass, 1984. (A fundamental work in the history of Indian inference.)
- Potter, K.H.** (1977) 'Introduction to the Philosophy of Nyāya-Vaiśeṣika', in *Encyclopedia of Indian Philosophies*, vol. 2, New Delhi: Motilal Banarsidass. (Chapter 9 of Part 1 gives a clear thirty-page survey of the Nyāya theory of inference.)
- Śankarasvāmin** (c.525) *Nyāyapraveśa*, trans. M. Tachikawa, *Journal of Indian Philosophy* 1: 111-45. (A clear and concise manual of logic for Buddhists.)
- Staal, F.** (1973) 'The Concept of *Pakṣa* in Indian Logic', *Journal of Indian Philosophy* 2: 156-66. (Useful and clear discussion of the difficulties of Indian inferential terminology.)
- Stcherbatsky, T.** (1931) *Buddhist Logic*, New York: Dover, 2 vols. (The first volume is a comprehensive introduction to Dharmakīrti's theory of knowledge and the second volume is a translation of Dharmakīrti's *Nyāyabindu*, together with the commentary by Dharmottara. The work is a remarkable achievement by one of the truly great scholars of the twentieth century. However, it is very difficult and somewhat dated.)
- Uddyotakara** (c.600) *Nyāyavārttika*, trans. G. Jhā, Delhi: Motilal Banarsidass, 1984. (Sets down one of the most important commentaries on the *Nyāyasūtra*.)
- Upaniṣads* (8th-6th century BC), *The Upaniṣads*, The Sacred Books of the East 1, 15, trans. F. Max Müller, Oxford: Clarendon Press, 1879-84; repr. New York: Dover, 1962. (These texts record views pertaining to a variety of matters, including cosmological, soteriological and epistemological matters, and are ascribed to thinkers thought to have lived between 1000 and 500 BC.)
- Vātsyāyana** (c.450) *Nyāyabhāṣya*, trans. G. Jhā, Delhi: Motilal Banarsidass, 1984. (The author sets down the earliest extant commentary on the *Nyāyasūtra*.)
- Vaiśeṣikasūtra* (1st century BC), ed. and trans. A.E. Gough, *The Vaiśeṣika Aphorisms of Kaṇāda*, Varanasi: E.J. Lazarus, 1873; repr. Delhi: Munshiram Manoharlal, 1975. (The earliest extant work on speculative ontology within the Brahmanical tradition.)
- Vidyabhusana, S.C.** (1921) *A History of Indian Logic*, Delhi: Motilal Banarsidass, 1971, 2 vols. (This is the standard, though dated, reference on the history of Indian logic.)
- Viśvanātha Pañcanana** (c.1650) *Bhāṣāpariccheda*, trans. Mādhavānanda, Almora: Advaita Ashrama, 1940. (A widely used manual of logic for the Nyāya school.)
- Witzel, M.** (1987) 'The Case of the Shattered Head', *Studien zur Indologie und Iranistik* 13/14: 363-415. (This article describes public contests which are thought to be the precursors of public debate.)
- Yaśovijaya** (c.1650) *Jaina Tarkabhāṣā*, trans. D. Bhargava, Delhi: Motilal Banarsidass, 1973. (A manual of logic

for Jainas.)

Infinitary logics

An infinitary logic arises from ordinary first-order logic when one or more of its finitary properties is allowed to become infinite, for example, by admitting infinitely long formulas or infinitely long or infinitely branched proof figures. The need to extend first-order logic became pressing in the late 1950s when it was realized that many of the fundamental notions of mathematics cannot be expressed in first-order logic in a way that would allow for their logical analysis. Because infinitary logics often do not suffer the same limitation, they have become an essential tool in mathematical logic.

Traditionally, logic - considered as the analysis of valid human reasoning, or something intimately connected with it - was seen as having to be finitary in its expressions, inferences and proofs, owing to the finitude of the human mind. This situation began to change in the second half of the nineteenth century, for two principal reasons. First, in philosophy, the Neo-Kantian conception of ‘*Geltung*’ (roughly, ‘validity’) gave rise both to a logical reconstructionist view in the philosophy of science - represented most prominently by Carnap - and to an anti-psychologistic strain of logical theory - represented most prominently by Frege. Second, in mathematics, although increased attention to foundations led to the elimination from analysis of a direct appeal to the infinite, it also led to increased emphasis on the infinite through the development of set theory.

In Löwenheim’s proof of his famous theorem (1915), we find perhaps the first deliberate use of infinitely long formulas and strings of quantifiers (see [Löwenheim-Skolem theorems and nonstandard models](#)). But development was interrupted for some thirty years by the influence of Hilbert, who stressed the finitary standpoint and focused on (the proof theory of) first-order logic and elementary recursion theory (see [Hilbert’s programme and formalism](#)).

In the early 1950s older modes of research (for example, investigation by logical reconstruction relying on infinitary means whenever necessary) were knit together again and model theory became fruitful. However, with the success of first-order model theory the deficiencies of this language for the expression of mathematical concepts (for example, compactness arguments) also became apparent. Thus, by the end of the decade, an extended model theory which searched for stronger logics began to emerge. Infinitary logics (ILs), in addition to fragments of n th-order logic, new quantifiers, and so on, proved to be a useful tool in increasing the expressive power of logical language.

Karp, Henkin, Tarski and others pioneered the development of ILs. At roughly the same time, ILs were introduced into proof theory as a means of streamlining and extending known techniques for consistency proofs. Since the Lindström theorems (see [Lindström 1969](#)) showed that interesting properties of first-order logic (for example, compactness; see [Hanf 1964](#)) must be lost if one tries to extend it, a new phase of research began when it was discovered that restriction to ‘admissible’ parts of a language not only preserves the usefulness of ILs but also yields suitable variants of lost properties (for example, Barwise compactness; see [Barwise 1975](#)).

Let $\mathcal{L}_{\kappa\lambda}$, with κ, λ (regular) infinite cardinals, denote an otherwise first-order language that allows for disjunctions (conjunctions) with fewer than κ disjuncts (conjuncts) and fewer than λ free variables (and hence fewer than λ quantifiers). (If κ or λ is arbitrary, we write ∞ instead.) Recall that ω is the cardinality of the natural numbers and ω_1 the least cardinal bigger than ω , that is, the cardinality of the reals in case the continuum hypothesis is true. In this notation ordinary first-order language $\mathcal{L}^1$, allowing for only finitely many disjuncts and free variables, becomes the finitary $\mathcal{L}_{\omega\omega}$.

Suppose we supplement a first-order language $\mathcal{L}_{\omega\omega}$ with an infinitary disjunctive, $\bigvee$, and add the formation rule: if Φ is a countable set of first-order formulas $\{\phi_n: n \in \omega\}$, then $\bigvee \Phi$ is a formula of $\mathcal{L}_{\omega_1\omega}$, being the countable disjunction of all $\phi \in \Phi$. Similarly for the infinitary conjunctive $\bigwedge \Phi$, though it might also be defined by De Morgan’s law as $\neg \bigvee \{\neg \phi_n: n \in \omega\}$.

To obtain a semantics for $\mathcal{L}_{\omega_1\omega}$, add to the standard semantics the clause

$F \models \bigvee \Phi$ iff there is some $\phi \in \Phi$ such that $F \models \phi$

and the corresponding clause for $\bigwedge \Phi$. Complementing the deduction rules of $\mathcal{L}_{\omega\omega}$ with the obvious variants for

$\bigvee$ (respectively, $\bigwedge$) results in a sound and complete calculus for $\mathcal{L}_{\omega_1\omega}$. Thus, for example, one might add the following rules:

- (1) $\vdash \phi \rightarrow \bigvee \Phi$, for $\phi \in \Phi$
- (2) If $\vdash \phi \rightarrow \psi$, for all $\phi \in \Phi$, then $\vdash \bigvee \Phi \rightarrow \psi$.

Two cautionary notes are in order here. First, completeness holds only with respect to the valid sentences and does not generally hold with regard to consequence (specifically, if the set of premises is not countable; see Scott 1965). Second, (2) shows that infinitely long derivations may occur, transforming the formal system into a semi-formal one.

In $\mathcal{L}_{\omega_1\omega}$ mathematical concepts that are not axiomatizable in ordinary $\mathcal{L}_{\omega\omega}$ become so. Simple examples include: the class of torsion groups, with the addition to the group axioms of the axiom

$$\forall x \bigvee \{ \underbrace{x \circ \dots \circ x}_{n \text{ times}} = e : 1 \leq n \in \omega \}$$

(where e denotes the neutral element of a group); the class of fields with characteristic unequal zero, with the addition to the field axioms of

$$\forall x \bigvee \{ \underbrace{1 + \dots + 1}_{n \text{ times}} = 0 : 2 \leq n \in \omega \};$$

and the class of structures isomorphic to the natural numbers, with the addition to the first two Peano axioms of

$$\forall x \bigvee \{ \underbrace{S \dots S}_n 0 = x : 0 \leq n \in \omega \}$$

(where S denotes the successor function). Among the logical metatheorems, downward Löwenheim-Skolem holds. Thus, if a sentence $\phi \in \mathcal{L}_{\omega_1\omega}$ is satisfiable, then it is countably satisfiable. The upward theorem is more difficult, requiring the use of the Hanf number $\mathbf{H}_\alpha$. If ϕ is satisfiable in all cardinalities $\mathbf{H}_\alpha$, with $\alpha < \omega_1$ and $\mathbf{H}$ a kind of superexponentiation of the alephs, then ϕ is satisfiable in all infinite cardinalities. Among other metatheorems that hold, Lopez-Escobar (1965), using a complete Gentzen system, was able to establish Craig's interpolation theorem and Beth's definability theorem (see also Karp 1964, Keisler 1971; Beth's theorem and Craig's theorem).

For languages between $\mathcal{L}_{\omega_1\omega}$ and $\mathcal{L}_{\infty\infty}$, as expected, the expressive power increases as the language becomes more infinitary. Thus, for example, each well-ordering is characterizable up to isomorphism by a sentence ϕ of $\mathcal{L}_{\infty\omega}$ and the notion of well-ordering itself becomes definable in $\mathcal{L}_{\omega_1\omega_1}$. Large languages thus admit more applications to mathematics (especially in algebraic subjects), even though many fundamental second-order concepts (for example, the class of topological spaces) remain inexpressible in such a language as $\mathcal{L}_{\infty\infty}$. In addition, as the language becomes larger, the rather coherent framework of metatheoretic results that still hold for $\mathcal{L}_{\omega_1\omega}$ dissolves (see Dickmann 1975, 1985).

In proof theory, the Gentzen-type consistency proofs of the early 1950s (see Proof theory) using ω -logic gained considerably in clarity by being reformulated in $\mathcal{L}_{\omega_1\omega}$ (see Tait 1968). Another kind of IL, ordinary $\mathcal{L}^1$ with infinitely long terms, was used to define functionals of finite type and was first employed by Tait (1965) to recast Gödel's *Dialectica* interpretation (producing a second link between infinitary logic and recursion theory in addition to that forged by the introduction of admissible sets; see Intuitionism §6).

See also: Logical and mathematical terms, glossary of

BERND BULDT

References and further reading

- Barwise, J.** (1975) *Admissible Sets and Structures: An Approach to Definability Theory*, New York: Springer. (Reference work concerning infinitary logic and admissible set theory; accessible.)
- Barwise, J.** (1980) 'Infinitary Logics', in E. Agazzi (ed.) *Modern Logic: A Survey*, Dordrecht: Reidel, 93- 112. (A mostly historical survey, which complements Moore (1990); very readable.)
- Barwise, J.** (1981) 'The Role of the Omitting Types Theorem in Infinitary Logic', *Archiv für mathematische Logik und Grundlagenforschung* 21: 55-68. (Surveys a metatheorem not treated here; useful.)

- Barwise, J. and Feferman, S.** (eds) (1985) *Model-Theoretic Logics*, New York: Springer, part A, chaps 2 and 3, and part C. (These first-rate surveys give ample information on infinitary logics in model theory, including a bibliography; advanced.)
- Dickmann, M.A.** (1975) *Large Infinitary Languages: Model Theory*, Amsterdam: North Holland. (Textbook; advanced.)
- Dickmann, M.A.** (1985) 'Larger Infinitary Languages', in J. Barwise and S. Feferman (eds) *Model-Theoretic Logics*, New York: Springer, 1985, 317-63. (Valuable survey; advanced.)
- Ebbinghaus, H.-D. and Flum, J.** (1995) *Finite Model Theory*, Berlin: Springer. (Textbook which also covers the use of infinitary logics in the growing field of finite model theory; very readable.)
- Ebbinghaus, H.-D., Flum, J. and Thomas, W.** (1984) *Mathematische Logik*, Heidelberg, Berlin and Oxford: Springer, 4th revised edn, 1996; trans. *Mathematical Logic*, Berlin: Springer, 2nd edn, 1994. (Introductory textbook including some infinitary logic and proofs of the Lindström theorems; very readable.)
- Feferman, S.** (1977) 'Theories of Finite Type Related to Mathematical Practice', in J. Barwise (ed.) *Handbook of Mathematical Logic*, Amsterdam: North Holland, ch. D4. (§8.3 provides some information on infinite terms; advanced.)
- Hanf, W.** (1964) 'Incompactness in Languages With Infinitely Long Expressions', *Fundamenta Mathematicae* 53: 309-24. (A milestone proving that ordinary compactness is lost in infinitary logics; advanced.)
- Henkin, L.A.** (1961) 'Some Remarks on Infinitely Long Formulas', in *Infinitistic Methods: Proceedings of the Symposium on Foundations of Mathematics*, New York: Pergamon Press, 167-83. (One of the earliest papers on the subject; readable.)
- Hugly, P. and Sayward, C.** (1983) 'Can a Language Have Indenumerably Many Expressions?', *History and Philosophy of Logic* 4: 73-82. (The authors claim not.)
- Karp, C.** (1964) *Languages With Expressions of Infinite Length*, Amsterdam: North Holland. (The first textbook on the subject; proceeds algebraically. Austerely written.)
- Keisler, H.J.** (1971) *Model Theory for Infinitary Logic: Logic With Countable Conjunctions and Finite Quantifiers*, Amsterdam: North Holland. (Proceeds along familiar model theoretic lines; very readable.)
- Lindström, P.** (1969) 'On Extensions of Elementary Logic', *Theoria* 35: 1-11. (The paper in which the two famous Lindström theorems are proved; advanced.)
- Lopez-Escobar, E.G.K.** (1965) 'An Interpolation Theorem for Denumerably Long Formulas', *Fundamenta Mathematicae* 57: 253-72. (Accessible.)
- Moore, G.H.** (1990) 'Proof and the Infinite', *Interchange* 21 (2): 46-60. (Survey of the history of infinitary logic from Aristotle to the early 1960s, which complements Barwise (1980); recommended.)
- Pohlers, W.** (1989) *Proof Theory: An Introduction*, Berlin: Springer. (Textbook using infinitary languages to obtain Gentzen-style consistency proofs; readable.)
- Scott, D.** (1965) 'Logic With Denumerably Long Formulas and Finite Strings of Quantifiers', in J.W. Addison et al. (eds) *The Theory of Models*, Amsterdam: North Holland, 329-41. (Influential paper arguing convincingly to the effect that $\mathcal{L}_{\omega_1\omega}$ is the 'truest' first-order extension, whereas $\mathcal{L}_{\omega_1\omega_1}$ is already second-order logic in disguise.)
- Scott, D. and Tarski, A.** (1958) 'The Sentential Calculus With Infinitely Long Expressions', *Colloquium Mathematicum* 6: 164-70. (One of the earliest papers on the subject; readable.)
- Tait, W.W.** (1965) 'Infinitely Long Terms of Transfinite Type', in J.N. Crossley and M.A.E. Dummett (eds) *Formal Systems and Recursive Functions*, Amsterdam: North Holland, 176-85. (Advanced.)
- Tait, W.W.** (1968) 'Normal Derivability in Classical Logic', in J. Barwise (ed.) *The Syntax and Semantics of Infinitary Languages*, New York: Springer, 204-36. (Milestone introducing $\mathcal{L}_{\omega_1\omega}$ to proof theory; advanced.)
- Tarski, A.** (1958) 'Remarks on Predicate Logic With Infinitely Long Expressions', *Colloquium Mathematicum* 6: 171-6. (One of the earliest papers on the subject; readable. Note 5 gives some historical remarks.)

Infinity

The infinite is standardly conceived as that which is endless, unlimited, immeasurable. It also has theological connotations of absoluteness and perfection. From the dawn of civilization, it has held a special fascination: people have been captivated by the boundlessness of space and time, by the mystery of numbers going on forever, by the paradoxes of endless divisibility and by the riddles of divine perfection.

The infinite is of profound importance to mathematics. Nevertheless, the relationship between the two has been a curiously ambivalent one. It is clear that mathematics in some sense presupposes the infinite, for instance in the fact that there is no largest integer. But the idea that the infinite should itself be an object of mathematical study has time and again been subjected to ridicule. In the nineteenth century this orthodoxy was challenged, with the advent of 'transfinite arithmetic'. Many, however, have remained sceptical, believing that the infinite is inherently beyond our grasp.

Perhaps their scepticism should be trained on the infinite itself: perhaps the concept is ultimately incoherent. It is certainly riddled with paradoxes. Yet we cannot simply jettison it. This is why the paradoxes are so acute. The roots of these paradoxes lie in our own finitude: it is self-conscious awareness of that finitude which gives us our initial sense of a contrasting infinite, and, at the same time, makes us despair of knowing anything about it, or having any kind of grasp of it. This creates a tension. We feel pressure to acknowledge the infinite, and we feel pressure not to. In trying to come to terms with the infinite, we are trying to come to terms with a basic conflict in ourselves.

1 Early Greek thought

The Greek word *peras* is usually translated as 'limit' or 'bound'. *To apeiron* denotes that which has no *peras*, the unlimited or unbounded: the infinite. *To apeiron* made its first significant appearance in early Greek thought with Anaximander of Miletus in the sixth century BC (see [Anaximander §2](#)). He thought of it as the boundless, imperishable, ultimate source of everything that is. He also thought of it as that to which all things must eventually return in order to atone for the injustices and disharmony which result from their transitory existence.

Anaximander was something of an exception, however. On the whole, the Greeks abhorred the infinite (as the old adage has it). More typical of that era were the Pythagoreans, a religious society founded by Pythagoras (see [Pythagoreanism §2](#)). They believed in two basic cosmological principles, *Peras* and *Apeiron*, the former subsuming all that was good, the latter all that was bad. They held further that the whole of creation was to be understood in terms of, and indeed was ultimately constituted by, the positive integers 1, 2, 3, ...; and that this was made possible by the fact that *Peras* was continuously subjugating *Apeiron* (the integers themselves, of course, are each finite). The Pythagoreans were followed to some extent in these beliefs by Plato, who also held that it was the imposition of limits on the unlimited that accounted for all the numerically definable phenomena that surround us.

However, the Pythagoreans soon learned to their dismay that they could not simply relegate the infinite to the role of cosmic villain. This was because of Pythagoras' own discovery that the square on the hypotenuse of a right-angled triangle is equal to the sum of the squares on the other two sides. Given this theorem, the ratio of a square's diagonal to each side is $\sqrt{2} : 1$. There are some good approximations to this ratio: it is a little more than 7 : 5, for example, and a little less than 17 : 12. Indeed there are approximations of any desired degree of accuracy. Nevertheless, given the basic tenets of Pythagoreanism, it ought to be *exactly* $p : q$, for some pair of positive integers p and q . The problem was that they discovered a proof that it is not, which they regarded as nothing short of catastrophic. According to legend, one of them was shipwrecked at sea for revealing the discovery to their enemies. The Pythagoreans had stumbled across the 'irrational' within mathematics. They had seen the limitations of the positive integers, and had thereby been forced to acknowledge the infinite in their very midst.

At around the same time, Zeno of Elea was formulating various celebrated paradoxes connected with the infinite (see [Zeno of Elea §6](#)). Best known of these is the paradox of Achilles and the tortoise: Achilles, who runs much faster than the tortoise, cannot overtake it in a race if he lets it start a certain distance ahead of him. For in order to do so he must first reach the point from which the tortoise started, by which time the tortoise will have advanced a fraction of the distance initially separating them; he must then make up this new distance, by which time the

tortoise will have advanced again; and so on *ad infinitum*. Such paradoxes, as well as having a profound impact on the history of thought about infinity, did much to reinforce early Greek hostility to the concept.

2 Aristotle

Aristotle's understanding of the infinite was an essentially modern one in so far as he defined it as the untraversable or never-ending. But he perceived a basic dilemma. On the one hand Zeno's paradoxes, along with a host of other considerations, show that the concept of the infinite really does resist a certain kind of application to reality. On the other hand there seems to be no prospect of doing without the concept, as the Pythagoreans had effectively realized. As well as $\sqrt{2}$, time seems to be infinite, numbers seem to go on *ad infinitum*, and space, time and matter all seem to be infinitely divisible.

Aristotle's solution to this dilemma was masterly. It has dominated all subsequent thought on the infinite, and until very recently was adopted by almost everyone who considered the topic. Aristotle distinguished between the 'actual infinite' and the 'potential infinite'. The actual infinite is that whose infinitude exists, or is given, at some point in time. The potential infinite is that whose infinitude exists, or is given, *over time*. All objections to the infinite, Aristotle insisted, are objections to the actual infinite. The potential infinite is a fundamental feature of reality. It is there to be acknowledged in any process which can never end: in the process of counting, for example, in various processes of division, or in the passage of time itself. The reason why paradoxes such as Zeno's arise is that we pay insufficient heed to this distinction. Having seen, for example, that there can be no end to the process of dividing a given racecourse, we somehow imagine that all those possible future divisions are already in effect there. We come to view the racecourse as already divided into infinitely many parts, and it is easy then for the paradoxes to take hold.

Even those later thinkers who did not share Aristotle's animosity towards the actual infinite tended to recognize the importance of his distinction. Often, though, Aristotle's reference to time was taken as a metaphor for something deeper and more abstract. This in turn usually proved to be something grammatical. Thus certain medieval thinkers distinguished between categorematic and syncategorematic uses of the word 'infinite'. Putting it very roughly, to use the word categorematically is to say that there is something with a property that surpasses any finite measure; to use the word syncategorematically is to say that, given any finite measure, there is something with a property that surpasses it. In the former case the infinite has to be instantiated 'all at once'. In the latter case it does not.

The categorematic-syncategorematic distinction heralds another distinction, whose importance to the infinite is hard to exaggerate. This is the distinction between saying that there is something of kind *X* to which each thing of kind *Y* stands in relation *R*, and saying that each thing of kind *Y* stands in relation *R* to something of kind *X* (not necessarily the same thing each time). This is referred to below as the 'Scope Distinction' (see [Scope](#)).

But Aristotle himself was not thinking in these very abstract terms. He took the references to time in his own account of the actual-potential distinction quite literally, and this gave rise to his most serious difficulty. He held that time (unlike space) is infinite. He also held that time involves constant activity, as exemplified in the revolution of the heavens. When our attention is focused on the future, there is no obvious problem with this. Past revolutions, however, because they are past, seem to have an infinitude which is by now completely given to us, and hence which is actual. This difficulty, in various different guises, has been a continual aggravation for philosophers who have wanted to see the infinite in broadly Aristotelian terms.

3 The rationalists and the empiricists

For over two thousand years Aristotle's conception of the infinite was regarded as orthodoxy. Often this conception was motivated by a kind of empiricism: the actual infinite was spurned on the grounds that we can never encounter it in experience. But does the potential infinite fare any better in this respect? Is experience of an infinitude that is given over time any less problematic than experience of an infinitude that is given all at once? The more extreme of the British empiricists were hostile to the infinite in all its guises. Where Aristotle had felt able to accept that space and time were infinitely divisible, [Berkeley](#) and [Hume](#) denied even that. They thought that the concept of the infinite was one that we could, and should, do without (see [Empiricism](#)).

This was partly a backlash against their rationalist predecessors. The rationalists had argued that we could form an

idea of the infinite, even though we could neither experience it nor imagine it. They thought that this idea was an innate one, and that it constituted, or helped to constitute, a vital insight into reality. They did not see any difficulty in this view. As Descartes put it, the fact that we cannot grasp the infinite does not preclude our touching it with our thoughts, any more than the fact that we cannot grasp a mountain precludes our touching it (see [Rationalism §2](#)).

Descartes believed that our idea of the infinite had been implanted in our minds by God (see [Descartes, R. §6](#)). Indeed this was the basis of one of his proofs of God's existence. Only a truly infinite being, Descartes argued, *could* have implanted such an idea in our minds. Note here the assimilation of the infinite to the divine: this was a legacy of medieval thought which is nowadays quite commonplace. But when the assimilation was first made, at the end of antiquity - most famously, by the Neo-Platonist Plotinus-it marked something of a turning point in the history of thought about the infinite (see [Plotinus §§3, 4](#)). Until then there had been a tendency to hear 'infinite' as a derogatory term. Henceforth, it was quite the opposite.

The empiricists, meanwhile, needed to defend their rejection of the infinite against the charge that it invalidated contemporary mathematics. They had more or less sophisticated ways of doing this, though in the case of geometry, where the problem was at its most acute, Hume took the rather cavalier step of simply denying certain crucial principles which mathematicians took for granted. (Berkeley's chief concern was with the use of infinitesimals in the recently invented calculus. In fact, his reservations were perfectly justified: it was a century before they were properly addressed.)

4 Kant

Kant played his characteristic role of conciliator in the debate on the infinite (see [Kant §§2, 5, 8](#)). He had an empiricist scepticism about the infinite, based on the fact that we cannot directly experience it. Nevertheless, he sided with the rationalists by insisting that there are certain formal or structural features of what we experience, which are accessible a priori and which *do* involve the infinite. Thus he thought that space and time were infinite (both in the sense of being infinitely extended and in the sense of being infinitely divisible): it is written into the form of whatever we experience that there can also be experience of how things are further out, further in, earlier or later. These, on Kant's view, were mathematical truths, a priori and unassailable.

But there is a question about how the topology of space and time *can* be a priori. Kant's celebrated reply was that space and time are not features of 'things in themselves'; they are part of an a priori framework which we contribute to our experience of things. What then of the *contents* of space and time, the physical universe as a whole? This was different. Kant did not think that what was physical was constructed a priori. Nor, on the other hand, did he think that it was ultimately real, that is to say real in a way that transcends any possible access we have to it. It had no features, on Kant's view, that exceed what we are capable of grasping through experience. So here the concept of the infinite did resist application. It still had what Kant regarded as a legitimate *regulative* use. That is, we could proceed *as if* the physical universe as a whole were infinite, thereby encouraging ourselves never to give up in our explorations. But we ultimately had no way of making sense of such infinitude. Kant was forced to take an extreme empiricist line by denying that the physical universe as a whole is infinitely big, that it has infinitely many parts and, going this time beyond Aristotle (thus bypassing the difficulty that had beset Aristotle himself), that it is infinitely old.

However, there was a dilemma. Kant was also forced to deny that the physical universe is *finite* in each of these three respects. Apart from anything else, to postulate infinite, empty space or time beyond the confines of the physical universe is itself to postulate that which exceeds what we are capable of grasping through experience.

This dilemma looks acute. Kant himself presented it in the form of a pair of 'antinomies'. These antinomies consisted of the principal arguments against the physical universe's being infinite in each of the specified respects, and the principal arguments against its being finite. But he believed that the dilemma contained the seeds of its own solution. If what is physical is not ultimately real - if there is no more to it than what we are capable of experiencing - then we are at liberty to deny that there is any such thing as the physical universe as a whole. There are only the finite physical things that are accessible to us through experience. The physical universe as a whole is neither infinite nor finite. It does not exist.

Kant's solution involved him in a direct application of the Scope Distinction (see §2 above). On the one hand he

affirmed that any finite physical thing is contained within something physical. On the other hand he denied that there is something physical within which any finite physical thing is contained. Both of these, the affirmation and the denial, were grounded in the fact that there is nothing we can identify in space and time such that we cannot identify more. This is fundamentally a fact about us: the fact that we are finite. Our identifications are always incomplete. What Kant added, in an idealist vein, was that what we cannot identify does not exist. Here, as in so many other places, we see how deeply involved with human finitude Kant's philosophy was, and how seriously he took it.

5 Post-Kantian metaphysics of the infinite

Metaphysical thought about the infinite since Kant has continued to be just as deeply involved with human finitude. Existentialists in particular have been greatly exercised by it, especially in its guise of mortality. But they have also for the most part recognized an element of the infinite within us. This too is Kantian. Kant believed that we are free rational agents, and that when our agency is properly exercised, it has an unconditioned autonomy that bears all the hallmarks of the truly infinite. For Kant, this was something which exalts us. But for many of the existentialists, still preoccupied with the fundamental fact of human finitude, it is something which is responsible for the deepest tensions within us, and thus for the absurdity of human existence (see [Existentialism §2](#)).

Hegel agreed with Kant that the truly infinite is to be found in the free exercise of reason (see [Hegel, G.W.F. §8](#)). But he took this further than Kant. He argued that reason is the infinite ground of everything. Everything that happens, on Hegel's view, can be understood as the activity of a kind of world-spirit, and this spirit *is* reason.

This led Hegel to a very non-Aristotelian conception of the infinite. For Hegel, the infinite was the complete, the whole, the unified. Aristotle's conception of the infinite as the never-ending was in Hegel's view quite wrong. He explained this conception as arising from our finite attempts to assimilate the truly infinite. And he described Aristotelian infinity as a 'spurious', or 'bad', infinity - a mere succession of finite elements, each bounded by the next, but never complete and never properly held together in unity. Such 'infinite' seemed to Hegel at turns nightmarish, then bizarre, then simply tedious, but always a pale, inadequate reflection of the truly infinite.

6 Modern mathematics of the infinite

Despite Kant's influence on Hegel, and despite his own commitment to infinite reason (as well as to infinite space), Kant certainly helped to propagate the Aristotelian tradition of treating the actual infinite with hostility and suspicion. As this tradition prevailed, the actual infinite came increasingly to be understood in the more general, non-temporal sense indicated in §2 above. Eventually, exception was being taken to any categorematic use of the word 'infinite'. The most serious challenge to this tradition, at least in a mathematical context, was not mounted until the nineteenth century, by Cantor, whose mathematical contribution to this topic is unsurpassed (see [Cantor, G.](#)).

Objections to the actual infinite had tended to be of two kinds. The first kind we have already seen: objections based on the fact that we can never encounter the actual infinite in experience. Objections of the second kind were based on the paradoxes to which the actual infinite gives rise. These paradoxes fall into two groups. The first group consists of Zeno's paradoxes and their variants. By the time Cantor was writing, however, the calculus (which had then reached full maturity) had done a great deal to mitigate these. Of more concern by then were the paradoxes in the second group, which had been known since medieval times. These were paradoxes of equinumerosity. They derive from the following principle: if (and only if) it is possible to pair off all the members of one set with all those of another, then the two sets must have just as many members as each other. For example, in a non-polygamous society, there must be just as many husbands as wives. This principle looks incontestable. However, if it is applied to infinite sets, it seems to flout Euclid's notion that the whole is greater than the part. For instance, it is possible to pair off all the positive integers with those that are even: 1 with 2, 2 with 4, 3 with 6 and so on.

Cantor accepted this principle. And, consistently with that, he accepted that there are just as many even positive integers as there are positive integers altogether. Far from being worried by this, he defined precisely what is going on in such cases, and then incorporated his definitions into a coherent, systematic and rigorous theory of the actual infinite, ready to be laid before any sceptical gaze.

It might be expected that, on this understanding, all infinite sets are the same size. (If they are, that is not unduly paradoxical.) But much of the revolutionary impact of Cantor's work came in his demonstration that they are not. There are different infinite sizes. This is a consequence of what is known as Cantor's theorem: no set, and in particular no infinite set, has as many members as it has subsets. In other words, no set is as big as the set of its subsets. If it were, then it would be possible to pair off all its members with all its subsets. But this is not possible. Suppose there *were* such a pairing and consider the set of members paired off with subsets not containing them. Whichever member was paired off with *this* subset would belong to it if and only if it did not belong to it (see [Cantor's theorem](#)).

In the course of developing these ideas, Cantor laid down some of the basic principles of the set theory which underlay them; he devised precise methods for measuring how big infinite sets are; and he formulated ways of calculating with these measures. In short, he established transfinite arithmetic.

Even so, there are many who remain suspicious of his work and who continue to think of the infinite in broadly Aristotelian terms. Cantor himself was forced to admit that there are some collections, including the collection of all things, which are so big that they cannot be assigned any determinate magnitude (their members cannot be given 'all at once'). Concerning such collections, he even sometimes said that they are 'truly' infinite. There is in fact a real irony here: Cantor's work can in many ways be regarded as corroborating Aristotelian orthodoxy.

Brouwer believed that Cantor had gone wrong in not showing sufficient respect for the first kind of objection to the actual infinite: that we cannot encounter it in experience. All Cantor had done, in Brouwer's view, was to demonstrate certain tricks that can be played with (finite) symbols, without addressing the question of how these tricks answer to experience. The relevant experience here - the experience to which any meaningful mathematical statement must answer, according to Brouwer and other members of his intuitionistic school - is our experience of time. It is by recognizing the possibility of separating time into parts, and then indefinitely repeating that operation over time, that we arrive at our idea of the infinite. And such infinitude is potential, not actual - in the most literal sense (see [Intuitionism §1](#)).

There was a very different critique of Cantor's ideas in the work of Wittgenstein, though it led to similar results (see [Wittgenstein §14](#)). Wittgenstein believed that insufficient attention had been paid (at least by those interpreting Cantor's work, if not by Cantor himself) to what he called the 'grammar' of the infinite, that is to certain fundamental constraints on what could count as a meaningful use of the vocabulary associated with infinity. In effect, Wittgenstein argued that the word 'infinite' could not be used categorically.

7 Human finitude

Problems about the infinite, we have seen, are grounded in our own finitude. On the one hand our finitude prevents us from being able to think of anything, including the whole of reality, as truly infinite. On the other hand it also prevents us from being able to think of anything finite - anything to that extent within our grasp - as the whole of reality. One way to reconcile these would be to deny that there is any such thing as the whole of reality and to argue that there are only bits of reality, each a part of some other. Here once again we see application of the Scope Distinction: every bit of reality is a part of something, but there is nothing of which every bit of reality is a part. Aristotle, Kant and even to an extent Cantor played out variations on this theme.

But one of the most pressing questions of philosophy still remains: in what exactly does our finitude consist? Some of the most striking features of that finitude are conditioned by our temporality. In particular, of course, there is the fact of our death. How are we to view death? Among the many subsidiary questions that this raises, there are two in particular which are superficially equivalent but between which it is important to distinguish. Putting them in the crudest possible terms (their refinement would be a large part of addressing them): (1) Is death a 'bad thing'? (2) Would immortality be preferable to mortality?

It can easily look as if these questions must receive the same answer. True, no sooner does one begin refining them than one sees all sorts of ways in which a full, qualified response to one can differ from a full, qualified response to the other. But it is in any case important to see how, even at this crude level, there is scope for answering 'yes' and 'no' respectively. Putting it very roughly, death is a bad thing because it closes off possibilities, but immortality would not be preferable to mortality because mortality is what gives life its most basic structure and, therewith, the possibility of meaning.

To answer ‘yes’ to (1) and ‘no’ to (2) in this way is once again to invoke the Scope Distinction. It is to affirm that at each time there is reason to carry on living for longer, while denying that there is reason to carry on living forever. Meaning, for self-conscious beings such as us, can extend further than any given limits. But it cannot extend further than them all.

If it is true that, in some sense, at some level and with all the myriad qualifications that are called for, the answer to (1) is ‘yes’ and the answer to (2) is ‘no’, then, coherent though that is, it points to a basic conflict in us: while it would not be good never to die, it is nevertheless never good to die. That conflict is one of the tragedies of human existence. It is also a version of the original conflict which underlies all our attempts to come to terms with the infinite. In thinking about the infinite, we are thinking, at a very deep level, about ourselves.

See also: [Continuum hypothesis, the §1](#); [Death](#)

A.W. MOORE

References and further reading

My thanks are due to Dartmouth publishers for permission to re-use material from the introduction to my book *Infinity*.

Aristotle (c. mid 4th century BC) *Physics*, Books III and IV, trans. E. Hussey, Oxford: Oxford University Press, 1983.(Book III, sects 4-8, presents the main elements of Aristotle’s account of the infinite.)

Benardete, J.A. (1964) *Infinity: An Essay in Metaphysics*, Oxford: Oxford University Press.(Entertaining and wide-ranging discussion of the infinite, with particular emphasis on its paradoxes.)

Bennett, J. (1971) ‘The Age and Size of the World’, in A.W. Moore (ed.) *Infinity*, Aldershot: Dartmouth, 1993. (Excellent discussion of Kant’s antinomies.)

Bolzano, B. (1851) *Paradoxes of the Infinite*, trans. D.A. Steele, ed. F. Prihonsky, London: Routledge & Kegan Paul, 1950.(Anticipates some of Cantor’s ideas, though with much less rigour. Historically significant.)

Brouwer, L.E.J. (1913) ‘Intuitionism and Formalism’, trans. A. Dresden, in P. Benacerraf and H. Putnam (eds) *Philosophy of Mathematics: Selected Readings*, Cambridge: Cambridge University Press, 1983.(Classic statement of some of the fundamental tenets of intuitionism.)

Cantor, G. (1895-7) *Contributions to the Founding of the Theory of Transfinite Numbers*, trans. P.E.B. Jourdain, New York: Dover, 1955.(Cantor’s second major publication, in which he establishes transfinite arithmetic. Of ground-breaking importance, but very technical.)

Hegel, G.W.F. (1812-16) *Science of Logic*, trans. A.V. Miller, London: Allen & Unwin, 1969.(Pages 116-57 and 225-38 provide the main elements of Hegel’s views.)

Heidegger, M. (1927) *Being and Time*, trans. J. Macquarrie and E. Robinson, Oxford: Blackwell, 1978.(*Locus classicus* of existential thought on human finitude, but exceedingly difficult.)

Hilbert, D. (1925) ‘On the Infinite’, trans. S. Bauer-Mengelberg, in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967.(A modern classic which defends ‘finitism’, a position according to which all references to the infinite are strictly meaningless though they can serve a useful function.)

Kant, I. (1781) *Critique of Pure Reason*, trans. N. Kemp Smith, London: Macmillan, 1933.(The section entitled ‘Transcendental Dialectic’ up to Book II, ch. 2 and omitting Book II, ch. 1 (A293/B349-A340/B398 and A405/B432-A567/B595) presents the antinomies and their solution.)

Lear, J. (1979-80) ‘Aristotelian Infinity’, in A.W. Moore (ed.) *Infinity*, Aldershot: Dartmouth, 1993.(Helpful and instructive account of Aristotle’s views, including a superb discussion of the problem of infinite past time.)

Moore, A.W. (1990) *The Infinite*, London: Routledge.(Introductory and partly historical study of all aspects of the infinite.)

Moore, A.W. (ed.) (1993) *Infinity*, Aldershot: Dartmouth.(Collection of the most important and influential articles on infinity published since 1950, with an extensive annotated bibliography and an introduction which expands on the material in this entry.)

Owen, G.E.L. (1957-8) ‘Zeno and the Mathematicians’, in A.W. Moore (ed.) *Infinity*, Aldershot: Dartmouth, 1993.(Thorough and scholarly discussion of Zeno’s paradoxes.)

Rucker, R. (1982) *Infinity and the Mind: The Science and Philosophy of the Infinite*, Sussex: Harvester Wheatsheaf.(Lively and fascinating account of the more mathematical aspects of the infinite. Defends a kind of mysticism.)

Russell, B. (1926) *Our Knowledge of the External World as a Field for Scientific Method in Philosophy*, London: Allen & Unwin.(Chapters 5-7 provide a vigorous defence of Cantor.)

- Salmon, W.C.** (ed.) (1970) *Zeno's Paradoxes*, Indianapolis, IN: Bobbs-Merrill. (Collection of some of the best known articles on Zeno's paradoxes.)
- Williams, B.** (1973) 'The Makropulos Case: Reflections on the Tedium of Immortality', in A.W. Moore (ed.) *Infinity*, Aldershot: Dartmouth, 1993. (Superb defence of the ambivalent attitude to death advocated in §7 above.)
- Wittgenstein, L.** (1975) *Philosophical Remarks*, trans. R. Hargreaves and R. White, ed. R. Rhees, Oxford: Blackwell. (Section XII and pages 304-14 provide a good representative sample of Wittgenstein's views.)

Intensional entities

Intensional entities are such things as concepts, propositions and properties. What makes them ‘intensional’ is that they violate the principle of extensionality; the principle that equivalence implies identity. For example, the concept of being a (well-formed) creature with a kidney and the concept of being a (well-formed) creature with a heart are equivalent in so far as they apply to the same things, but they are different concepts. Likewise, although the proposition that creatures with kidneys have kidneys and the proposition that creatures with hearts have kidneys are equivalent (both are true), they are not identical. Intensional entities are contrasted with extensional entities such as sets, which do satisfy the principle of extensionality. For example, the set of creatures with kidneys and the set of creatures with hearts are equivalent in so far as they have the same members and, accordingly, are identical. By this standard criterion, each of the following philosophically important types of entity is intensional: qualities, attributes, properties, relations, conditions, states, concepts, ideas, notions, propositions and thoughts.

All (or most) of these intensional entities have been classified at one time or another as kinds of universals. Accordingly, standard traditional views about the ontological status of universals carry over to intensional entities. Nominalists hold that they do not really exist. Conceptualists accept their existence but deem it to be mind-dependent. Realists hold that they are mind-independent. Ante rem realists hold that they exist independently of being true of anything; in re realists require that they be true of something.

1 History

Contemporary use of the term ‘intension’ derives from the traditional logical doctrine that an idea has both an extension and an intension (or comprehension). This doctrine is explicit in most modern logicians of the nineteenth century, implicit in many medieval logicians and, arguably, present in Porphyry and Aristotle. Although there is divergence in formulation, most of these thinkers accept that the extension of an idea (or concept) consists of the subjects to which the idea applies, and the intension consists of the attributes implied by the idea. The Port Royalists (see [Arnauld 1662](#)) tell us, ‘the comprehension of the idea of triangle includes extension, figure, three lines, three angles, and the equality of these three angles to two right angles, and so forth’. This yields the ‘law of inverse ratio’: the larger the intension, the smaller the extension. (Note that these traditional doctrines were expressed with the use of plurals and without explicit commitment to classes or sets.) If the extension consists of the individuals (as opposed to species) to which an idea applies, it is evident that two ideas could have the same extension but different intensions. This is the tie to Russell’s subsequent notion of an intensional propositional function (1910-13) and leads to our contemporary usage of ‘intensional entity’ and ‘extensional entity’.

In contemporary philosophy, it is linguistic expressions, rather than concepts, that are said to have intensions and extensions. The intension is the concept expressed by the expression, and the extension is the set of items to which the expression applies. On the standard picture, that set is the same as the set of items to which the concept applies, hence the standard view that intension determines extension. This usage resembles Mill’s use of ‘denotation’ and ‘connotation’ and Frege’s use of ‘*Bedeutung*’ and ‘*Sinn*’ (see [Mill, J.S. §§2-3](#); [Frege, G. §3](#)).

The systematic study of intensional entities has been pursued largely in the context of intensional logic; that part of logic in which the principle of substitutivity of equivalent expressions fails. For example, ‘Necessarily, creatures with kidneys have kidneys’ is true whereas ‘Necessarily, creatures with hearts have kidneys’ is false. Following Frege (1892), Whitehead and Russell (1910-13), Carnap (1947) and Church (1951), the now standard explanation of this failure of substitutivity is this: the truth-value of these sentences is determined, not (just) by the extension of ‘creature with kidney’ and ‘creature with heart’ (that is, the set of creatures with kidneys) but (also) by the intensions of these expressions, that is, by the intensional entities which they express (the concept of being a creature with a kidney and the concept of being a creature with a heart). Because these intensions are different, the sentences can have different truth-values.

This account yields a method for determining precise intensional identity conditions: intensional entities should be as finely discriminated as is necessary for explaining associated substitutivity failures in intensional logic. For example, this method has led most philosophers to hold that logically equivalent concepts can be distinct, and even that in a correct definition the definiens and the definiendum can express different concepts. This is not to say that the properties corresponding to those concepts would be distinct; on the contrary, the definition would be correct

only if the properties were identical. Such considerations lead to the view that there are also more coarsely grained intensional entities (properties, relations, conditions in the world). Independent arguments in metaphysics, epistemology, philosophy of science and aesthetics seem to support the same conclusion (see, for example, [Armstrong 1978](#), [Bealer 1982](#), [Lewis 1986](#)). On this view, properties and relations (as opposed to concepts) play a primary role in the non-arbitrary categorization and identification of objects, in the constitution of experience, in description and explanation of change, in the theory of inductive inference, in the analysis of objective similarities and in the statement of supervenience principles. In the resulting picture there are both fine-grained intensions, which play the role of cognitive and linguistic contents, and coarse-grained intensions, which play a constitutive role in the structure of the world.

Systematic theories of intensional entities have often incorporated some form of extensional reductionism. The leading reductions (possible worlds, propositional complexes, propositional functions) are critically reviewed below; non-reductionist approaches to intensional entities are then discussed. At issue is the question of what intensional entities *are*. Are they identical to extensional functions, ordered sets, sequences and so on? Or are they *sui generis* entities, belonging to an altogether new category?

2 Extensional reductions

On the possible worlds reduction (see [Lewis 1986](#); [Stalnaker 1984](#)), a proposition is either a set of possible worlds or a function from possible worlds to truth-values, and properties are functions from possible worlds to sets of possible (usually non-actual) objects (see [Semantics, possible worlds §9](#)). Many people find the possible worlds theory intuitively implausible. Are familiar sensible properties (for example, colours, shapes, aromas) really functions? When I am aware that I am in pain, is a set (function) of possible worlds really the object of my awareness? Besides this intuitive objection, there are familiar epistemological and metaphysical objections to a theory that is truly committed to the existence of things that do not exist actually. Finally, the possible worlds reduction implies that necessarily equivalent intensions are identical, but as seen above this is implausible in the case of concepts and propositions. Certain possible worlds theorists (for example, [Cresswell 1985](#)) have responded to this problem by holding that concepts and propositions are ordered sets (sequences, abstract trees) whose elements are possible worlds constructs. This revisionary view is a variation of the propositional-complex reduction.

According to this reduction (see [Perry and Barwise 1983](#)), concepts and propositions are nothing but ordered sets (sequences, abstract trees, partial functions) whose ultimate elements are properties, relations and perhaps individuals. For example, the proposition that you are running is the ordered set $\langle \text{running, you} \rangle$; the proposition that you are running and I am walking is $\langle \text{conjunction, } \langle \text{running, you} \rangle, \langle \text{walking, me} \rangle \rangle$;

and so forth. As with the possible worlds theory, this theory clashes with intuition. When I am aware that I am in pain, is an ordered set the object of my awareness? When I see that you are running, do I see an ordered set? Furthermore, there is in principle no way to determine which ordered set I allegedly see. Is it $\langle \text{running, you} \rangle$? Or is it $\langle \text{you, running} \rangle$? The choice seems utterly arbitrary, but if the reduction were correct, there would have to be a fact of the matter.

A rather different kind of difficulty arises in connection with ‘transmodal quantification’. Consider the following intuitively true sentence:

(1) Every x is such that, necessarily, for every y , it is either possible or impossible that $x = y$.

In symbols,

(2) $(\forall x)\Box(\forall y)(\text{Possible } [x = y] \vee \text{Impossible } [x = y])$.

By the propositional-complex theory, this is equivalent to

(3) $(\forall x)\Box(\forall y)(\text{Possible } \langle x, \text{identity}, y \rangle \vee \text{Impossible } \langle x, \text{identity}, y \rangle)$.

The singular term ‘ $\langle x, \text{identity}, y \rangle$ ’ may be thought of as a definite description, ‘the ordered set whose elements

are x , identity and y '. Accordingly, there are two ways to understand (3). On the first construal (corresponding to the narrow scope reading of the definite description), (3) implies

$$(\forall x) \Box (\forall y) (\exists v) v = \langle x, \text{identity}, y \rangle.$$

However, by the principle that, necessarily, a set exists only if its elements exist, this implies

$$(\forall x) \Box (\exists v) v = x.$$

That is, everything necessarily exists. A manifest falsehood. On the alternative construal of (3), corresponding to the wide scope reading of the definite description, 'the ordered set $\langle x, \text{identity}, y \rangle$ ', (3) implies that every x is such that, necessarily, for all y , there exists an actual set $\langle x, \text{identity}, y \rangle$. That is

$$(\forall x) \Box (\forall y) (\exists_{\text{actual}} v) v = \langle x, \text{identity}, y \rangle.$$

But, by the principle that, necessarily, a set is actual only if its elements are actual, this implies

$$\Box (\forall y) y \text{ is actual.}$$

That is, necessarily, everything (including everything that could exist) is already actual. Another manifest falsehood. So either way, (3) implies something false. But (3) is the propositional-complex theorists' way of representing the true sentence (2). So the propositional-complex theory appears unable to handle intuitively true sentences such as (2). (This style of argument can be exploited in a defence of *ante rem* realism.)

According to the propositional-function theory (see [Whitehead and Russell 1910-13](#)), a property is a function from objects to propositions, where propositions are taken to be primitive entities. For example, the property 'being red' is the function $(\lambda x)(\text{the proposition that } x \text{ is red})$. For any given object x , the proposition that x is red is the result of applying this function to the argument x . That is, the proposition that x is red is $(\lambda x)(\text{the proposition that } x \text{ is red})(x)$. But are the familiar sensible properties really functions? Once again, this is highly implausible and should be rejected on this ground alone if an acceptable alternative exists.

Besides this sort of intuitive difficulty, there are several technical difficulties. Consider an illustration involving properties of integers. Being even = being an x such that x is divisible by two; and being self-divisible = being an x such that x is divisible by x . Given the propositional-function theory, the following identities hold:

$$\begin{aligned} \text{that two is even} &= (\lambda x)(\text{that } x \text{ is even})(\text{two}) \\ &= (\lambda x)(\text{that } x \text{ is divisible by two})(\text{two}) \\ &= \text{that two is divisible by two} \\ &= (\lambda x)(\text{that } x \text{ is divisible by } x)(\text{two}) \\ &= (\lambda x)(\text{that } x \text{ is self-divisible})(\text{two}) \\ &= \text{that two is self-divisible.} \end{aligned}$$

However, that two is even and that two is self-divisible are plainly different propositions: certainly someone could be consciously and explicitly thinking the former while not consciously and explicitly thinking the latter. Indeed, someone who is thinking that two is even might never have considered the concept of self-divisibility. (Analogous difficulties arise for Church's somewhat similar propositional-function theory of concepts ([1951](#)). Incidentally, since properties are not propositional functions, using function abstracts to denote them invites confusion. A growing practice is to use $\lceil v: A \rceil$ to denote the property of being a v such that A , just as in set theory $\{v: A\}^1$ denotes the set of things v such that A .)

3 Non-reductionist approaches

Each of the preceding problems arises from the attempt to reduce intensional entities of one kind or another to extensional entities; either extensional functions or sets. These difficulties have led some theorists to adopt non-reductionist approaches (see [Bealer 1979, 1982](#)). Consider the following truisms. The proposition that $A \& B$ is the conjunction of the proposition that A and the proposition that B . The proposition that not A is the negation of the proposition that A . The proposition that Fx is the predication of the property F of x . The proposition that there

exists an F is the existential generalization of the property F , and so on. These truisms tell us what these propositions essentially are: they are by nature conjunctions, negations, singular predications, existential generalizations and so on. These are rudimentary facts which require no further explanation and for which no further explanation is possible. Until the advent of extensionalism, this was the standard view.

By adapting techniques developed in the algebraic tradition in extensional logic, one is able to develop this non-reductionistic approach. Examples such as those just given isolate fundamental logical operations: conjunction, negation, singular predication, existential generalization and so on. Intensional entities are then taken as *sui generis* entities; the aim is to analyse their behaviour with respect to the fundamental logical operations.

An intensional algebra is a structure consisting of a domain D , a set of logical operations and a set of possible extensionalization functions. The domain divides into subdomains: particulars, propositions, properties, binary relations, ternary relations and so on. The set of logical operations includes those listed above plus certain auxiliary operations. The possible extensionalization functions assign a possible extension to relevant items in the domain: each proposition is assigned a truth-value; each property is assigned a set of items in D ; each binary relation is assigned a set of ordered pairs of items in D , and so on. One extensionalization function is singled out as the actual extensionalization function: the propositions which are true relative to it are the propositions which are actually true, and so on. (More formally, an intensional algebra is a structure $\langle D, \tau, K, G \rangle$. D divides into subdomains: $D_{-1}, D_0, D_1, D_2, \dots$. D_{-1} consists of particulars; D_0 , propositions; D_1 , properties; D_2 , binary relations, and so on. τ is a set of logical operations on D . K is a set of extensionalization functions. G is a distinguished function in K which is the actual extensionalization function.)

To illustrate how this approach works, consider the operation of conjunction, conj . Let H be an extensionalization function. Then, conj satisfies the following: for all propositions p and q in D , $H(\text{conj}(p, q)) = \text{true}$ iff $H(p) = \text{true}$ and $H(q) = \text{true}$. Similarly, if neg is the operation of negation, then for all propositions p in D , $H(\text{neg}(p)) = \text{true}$ iff $H(p) = \text{false}$. Likewise, for singular predication pred_s , which takes properties F in D and items y in D to propositions in D : $H(\text{pred}_s(F, y)) = \text{true}$ iff y is in the extension $H(F)$.

This non-reductionistic approach can be extended to more complex settings in which, for example, both fine-grained intensional entities (concepts, propositions) and coarse-grained intensional entities (properties, relations, conditions) are treated concurrently and in which a relation of correspondence between the two types of intensional entities can be characterized in terms of the fundamental logical operations. The non-reductionist approach can thus accommodate the fine-grained intensional entities which serve as cognitive and linguistic contents and also the more coarsely grained intensional entities which play a constitutive role in the structure of the world.

See also: [Nominalism](#); [Intensional logics](#); [Logical and mathematical terms, glossary of](#); [Modal logic, philosophical issues in](#); [Second-order logic, philosophical issues in](#); [Property theory](#)

GEORGE BEALER

References and further reading

- Aczel, P.** (1980) 'Frege Structures and the Notions of Propositions, Truth and Set', in J. Keisler, J. Barwise and K. Kunen (eds) *The Kleene Symposium*, Amsterdam: North Holland, 31-59.(Development of the propositional-function view.)
- Aczel, P.** (1989) 'Algebraic Semantics for Intensional Logics, I', in G. Chierchia, B. Partee and R. Turner (eds) *Property Theories, Type Theories, and Semantics*, Dordrecht: Kluwer, 17-46.(Further development of the propositional-function view.)
- Armstrong, D.M.** (1978) *A Theory of Universals*, Cambridge: Cambridge University Press.(Study of coarse-grained intensional entities, discussed at close of §1.)
- Arnauld, A.** (1662) *Logic, or the Art of Thinking: The Port-Royal Logic*, trans. T.S. Baynes, Edinburgh: Sutherland & Knox, 1850.(Presentation of the traditional doctrine of intension and extension, discussed in §1.)
- Bealer, G.** (1979) 'Theories of Properties, Relations, and Propositions', *Journal of Philosophy* 76: 643-8. (Development of the non-reductionist view discussed in §3.)
- Bealer, G.** (1982) *Quality and Concept*, Oxford: Clarendon Press.(Expansion of the material in §1 and development of the non-reductionist view discussed in §3.)
- Bealer, G.** (1993a) 'Universals', *Journal of Philosophy* 90: 5-32.(Defence of the *ante rem* theory of intensional

- entities.)
- Bealer, G.** (1993b) 'A Solution to Frege's Puzzle', in J. Tomberlin (ed.) *Philosophical Perspectives*, Atascadero, CA: Ridgeview Press, vol. 7, 17-61.(Application of the non-reductionist view to some outstanding problems in the logic for intensional entities.)
- Bealer, G. and Mönnich, U.** (1989) 'Property Theories', in D. Gabbay and F. Guentner (eds) *Handbook of Philosophical Logic*, Dordrecht: Reidel, vol. 4, 133-257.
- Carnap, R.** (1947) *Meaning and Necessity*, Chicago, IL: University of Chicago Press.(Early presentation of the contemporary use of 'intension' and 'extension' and precursor to the possible worlds theory.)
- Church, A.** (1951) 'A Formulation of the Logic of Sense and Denotation', in P. Henle, H.H. Kallen, S.K. Langer (eds) *Structure, Method, and Meaning: Essays in Honor of Henry M. Scheffer*, New York: Liberal Arts Press, 3-24.(Development of a Fregean version of the propositional-function view.)
- Cresswell, M.J.** (1985) *Structured Meanings*, Cambridge, MA: MIT Press.(A combination of the possible worlds theory and the propositional-complex theory.)
- Frege, G.** (1892) 'Über Sinn und Bedeutung', *Zeitschrift für Philosophie und philosophische Kritik* 100: 25-50; trans. M. Black, 'On Sense and Meaning', in *Translations from the Philosophical Writings of Gottlob Frege*, ed. P.T. Geach and M. Black, Oxford: Blackwell, 3rd edn, 1980.(Defence of the theory that meaningful expressions typically have both a sense and a reference.)
- Frisch, J.C.** (1969) *Extension and Comprehension in Logic*, New York: Philosophical Library.
- Lewis, D.K.** (1986) *On the Plurality of Worlds*, Oxford: Blackwell.(Defence of coarse-grained intensions and the possible worlds theory.)
- Lewis, D.K.** (1983) 'New Work for a Theory of Universals', *Australasian Journal of Philosophy* 61: 343-77.
- Menzel, C.** (1986) 'A Complete Type-Free "Second-Order" Logic and its Philosophical Foundations', in *Report No. CSLI-86-40*, Stanford, CA: Center for the Study of Language and Information, Stanford University. (Development of the non-reductionist view discussed in §3.)
- Mönnich, U.** (1983) 'Toward a Calculus of Concepts as a Semantical Metalanguage', in R. Bäuerle, C. Schwarze and A. von Stechow (eds) *Meaning, Use, and Interpretation of Language*, Berlin: de Gruyter, 342-60.
- Parsons, T.** (1980) *Nonexistent Objects*, New Haven, CT: Yale University Press.(Development of the non-reductionist view in combination with a Meinongian theory of nonexistent objects.)
- Perry, J. and Barwise, J.** (1983) *Situations and Attitudes*, Cambridge, MA: MIT Press.(Development of the propositional-complex view.)
- Richard, M.** (1990) *Propositional Attitudes: An Essay on Thoughts and How We Ascribe Them*, New York: Cambridge University Press.
- Salmon, N.** (1986) *Frege's Puzzle*, Cambridge, MA: MIT Press.(Development of the propositional-complex view.)
- Soames, S.** (1987) 'Direct Reference, Propositional Attitudes, and Semantic Content', *Philosophical Topics* 15: 47-87.
- Stalnaker, R.** (1984) *Inquiry*, Cambridge, MA: MIT Press.(Defence of the possible worlds view.)
- Turner, R.** (1987) 'A Theory of Properties', *Journal of Symbolic Logic* 52: 455-72.
- Turner, R.** (1991) *Truth and Modality for Knowledge Representation*, Cambridge, MA: MIT Press.(Development of the propositional-function view.)
- Turner, R. and Chierchia, G.** (1988) 'Semantics and Property Theory', *Linguistics and Philosophy* 11: 261-302.
- Whitehead, A.N. and Russell, B.A.W.** (1910-13) *Principia Mathematica*, vol. 1, Cambridge: Cambridge University Press, 3 vols; 2nd edn, 1925-7.(A classic of mathematical logic, which includes a development of intensional logic based on the propositional-function theory.)
- Zalta, E.** (1983) *Abstract Objects: An Introduction to Axiomatic Metaphysics*, Dordrecht: Reidel.
- Zalta, E.** (1988) *Intensional Logic and the Metaphysics of Intentionality*, Cambridge, MA: MIT Press. (Development of the non-reductionist view in combination with a Meinongian theory of nonexistent objects.)

Intensional logics

Intensional logics are systems that distinguish an expression's intension (roughly, its sense or meaning) from its extension (reference, denotation). The purpose of bringing intensions into logic is to explain the logical behaviour of so-called intensional expressions. Intensional expressions create contexts which violate a cluster of standard principles of logic, the most notable of which is the law of substitution of identities - the law that from $a = b$ and $P(a)$ it follows that $P(b)$. For example, 'obviously' is intensional because the following instance of the law of substitution is invalid (at least on one reading): $\text{Scott} = \text{the author of Waverley}$; obviously $\text{Scott} = \text{Scott}$; so, obviously $\text{Scott} = \text{the author of Waverley}$. By providing an analysis of meaning, intensional logics attempt to explain the logical behaviour of expressions such as 'obviously'. On the assumption that it is intensions and not extensions which matter in intensional contexts, the failure of substitution and related anomalies can be understood.

Alonzo Church pioneered intensional logic, basing it on his theory of types. However, the widespread application of intensional logic to linguistics and philosophy began with the work of Richard Montague, who crafted a number of systems designed to capture the expressive power of natural languages. One important feature of Montague's work was the application of possible worlds semantics to the analysis of intensional logic. The most difficult problems concerning intensional logic concern the treatment of propositional attitude verbs, such as 'believes', 'desires' and 'knows'. Such expressions pose difficulties for the possible worlds treatment, and have thus spawned alternative approaches.

1 Intension and extension

Gottlob Frege believed that expressions of language have both an extension (reference, denotation) and an intension (sense, meaning; see [Sense and reference](#)). He distinguished intensions from extensions to explain a failure of the law of substitution of identity. From 'The author of *Waverley* = Scott' and 'John admires the author of *Waverley*' it does not follow that 'John admires Scott', for John may think that Scott wrote *Ulysses*, which he hates. (Frege illustrated the point with 'the morning star' and 'the evening star'.) One way out of the puzzle is to assume that the second sentence reports a relation (of admiration) between John and the meaning (or intension) of '*Waverley*', rather than a relation between John and the extension of 'the author of *Waverley*', that is, Scott. The fact that the meanings of 'Scott' and 'the author of *Waverley*' differ can help explain why substitution fails.

The failure of substitution in the preceding example is caused by the word 'admires'. Such intensional expressions violate several other standard principles of logic. A notable case is existential generalization: $P(a)$; therefore there exists an x such that $P(x)$. Existential generalization fails because, for example, 'John admires Superman' does not entail that Superman actually exists. The stock of intensional expressions of English is very large. It includes phrases such as 'it is necessary that', 'it used to be the case that', 'believes that', 'it ought to be the case that', and many more.

Under a broad construal, intensional logic includes any logic that accounts for the deductive behaviour of one or more of the intensional expressions. So modal logic, tense logic and logics of belief and obligation all count as intensional logics from this point of view. However, the term 'intensional logic' more commonly refers to those systems in which some explicit symbolic device is introduced for distinguishing intensions from extensions. These are the systems described here.

2 Possible worlds semantics

Despite its failings, possible worlds semantics has come to play a central role in intensional logic (see [Possible worlds](#)). The basic idea goes back to Rudolf Carnap. Why do the meanings (or intensions) of 'Scott' and 'the author of *Waverley*' differ? Because there are possible worlds where the author of *Waverley* was not Scott. We can imagine a world, for example, where it was James Joyce or Charles Dickens who wrote *Waverley*. If we consider the extensions that 'the author of *Waverley*' takes across different possible worlds, we will see variations from one world to the next. It makes sense, then, to identify the *intension* of the expression with the pattern of its *extensions* across different possible worlds. Seen in this way, its intension is a function that takes each possible world into a corresponding extension. The value of the function is Scott in the real world, Joyce in another world, and Dickens

in another. The intension for ‘Scott’, on the other hand, is the constant function that has the value Scott in each possible world.

Intensions for sentences, predicates and expressions in other grammatical categories can be treated in a similar way. The extension of a sentence is its truth-value. The intension is identified with the pattern of truth-values the sentence takes across the different possible worlds. It is therefore a function that assigns to each possible world a corresponding truth-value. The same tactic can be used for predicates. The extension of the predicate ‘is blue’ is traditionally taken to be the set of objects that are blue. Its intension is a function that takes each of the possible worlds into a corresponding set of objects (the set of blue objects in that world).

3 Montague’s intensional logic

Alonzo Church (1951) was the pioneer in intensional logic, basing it on his theory of types, but Richard Montague’s work has been crucial to its widespread application in philosophy and linguistics (see Montague 1974). The system IL described here is a variant of Montague’s approach, due to Gallin (1975). Intensions and extensions for sentences, names, predicates, adjectives, prepositions and so on are different sorts of things. Therefore we need to distinguish a rich variety of ‘types’ (grammatical categories or parts of speech). Most intensional logics are constructed from some form of type theory, a higher-order logic that allows a completely general formation of types (see Theory of types; Second- and higher-order logics).

In IL there are two basic types: t for truth-values and e for entities (or objects). Extensions of sentences belong to type t and extensions of proper names belong to type e . An unlimited variety of complex types can be defined from the two basic types as follows. If α and β are types then (α, β) is another type, and if α is a type then so is (s, α) , where the letter ‘ s ’ is used to indicate the set of possible worlds (or situations). The type (α, β) is for the set of all possible functions which assign to each thing with type α a value with type β . For example, if α is the type for the extension of some expression, then (s, α) is the type for its intension.

Extensions and intensions for the different grammatical categories of English can be assigned one of these types in the following way. Extensions of sentences and proper names are of types t and e , respectively, and the types for their intensions are (s, t) and (s, e) , respectively. The type for predicate extensions is (e, t) : the type for functions that take entities into truth-values. The idea is that the extension for ‘is blue’ is a function that assigns the truth-value T (for true) to those entities that are blue and F (for false) to those that are not. This has the same effect as treating the extension of ‘is blue’ as a set of entities, but it simplifies the definition of the complex types. Extensions for transitive verbs should take a pair of entities (one for the subject and one for the object of the verb) into a truth-value, so transitive verb extensions are assigned to type $(e, (e, t))$. The types for intensions of intransitive and transitive verbs would be $(s, (e, t))$ and $(s, (e, (e, t)))$, respectively.

IL has both constants and variables in all the possible types. It also contains the logical symbols ‘ $\equiv$ ’, ‘ λ ’, ‘ $\wedge$ ’, ‘ $\vee$ ’, ‘ $[$ ’ and ‘ $]$ ’. These symbols allow the construction of complex terms (or expressions). The symbols ‘ $\wedge$ ’ and ‘ $\vee$ ’ (called ‘cap’ and ‘cup’) indicate the intension and extension of a term, respectively. So when A has type α , $\wedge A$ is the intension of A and must have type (s, α) . When A has type (s, α) , then $\vee A$ is A ’s extension and belongs to type α . The symbol ‘ $\equiv$ ’ indicates identity. When A and B are two terms of the same type, the expression ‘ $A \equiv B$ ’ belongs to the t (truth-value) type, with the understanding that $A \equiv B$ is true if and only if A and B name the same thing. $[A \ B]$ indicates the result of applying the function that A names to whatever B names. So when A belongs to type (α, β) and B to type α , $[A \ B]$ belongs to type β . For example, if A is ‘runs’ with type (e, t) and B is ‘John’ with type e , then $[runs \ John]$ has type t , and symbolizes the extension of the English sentence ‘John runs’. When x is a variable with type α , and A is an expression with type β , then $\lambda x A$ is an expression with type (α, β) , indicating functional abstraction. For example, when $A =$ ‘Mary loves x ’, $\lambda x A$ names the function that assigns to each entity o either T or F depending on whether Mary loves o or not. Hence, in this example, $\lambda x A$ corresponds to the predicate ‘is loved by Mary’.

IL has strong expressive powers. All of the symbols of predicate logic can be defined in it - including propositional connectives (negation, conjunction and so on) and quantifiers in all the types. Furthermore, the presence of ‘ $\wedge$ ’ ensures that an S5 strength modal operator ‘ $\Box$ ’ (read ‘it is necessary that’) can be defined (for a description of S5, see Modal logic). IL is on a firm technical footing. Gallin has devised rules and axioms for IL and has shown how to provide a (generalized) semantics for this system for which IL is both consistent and complete.

Intensional logics such as IL have been applied to tenses, necessity, obligation and many other intensional constructions. The same ideas can be used to analyse indexicals such as ‘I’, ‘it’ and ‘that’, with the understanding that possible worlds include contexts of utterance that help fix the extensions of these terms. A hint at how IL is applied to language can be given by showing how to explain the failure of substitution in the case of John’s admiring the author of *Waverley*. Assume that each of the words of English is assigned an appropriate type of IL. The verb ‘admires’ is intensional, meaning that the object of the verb is an *intension* of type (s, e) . So the extension of ‘admires’ falls in type $(e, ((s, e), t))$, not $(e, (e, t))$. It follows that ‘John admires Scott’ is represented in IL by $[[\text{admires John}] \wedge S]$, where S abbreviates ‘Scott’. The cap ‘ $\wedge$ ’ makes it clear that it is the intension of ‘Scott’ that is at issue. Similarly, $[[\text{admires John}] \wedge W]$ represents ‘John admires the author of *Waverley*’. The claim that Scott is the author of *Waverley* is written ‘ $S \equiv W$ ’. Since $S \equiv W$ only says that S and W have the same *extension*, the inference from $S \equiv W$ and $[[\text{admires John}] \wedge W]$ to $[[\text{admires John}] \wedge S]$ is invalid. An identity involving S and W does not warrant substitution of $\wedge S$ for $\wedge W$.

4 The problem of propositional attitudes

While intensional logics in the style of IL appear to work nicely in the case of tense, necessity, obligation, indexicals and the like, serious problems arise in certain cases when IL is applied to so-called propositional attitude verbs, such as ‘believes’, ‘wants’, ‘knows’ and ‘admires’ (see [Propositional attitude statements](#)). The fundamental difficulty is that systems based on IL validate substitution of *necessary* identities: if expressions A and B have the same extensions in every possible world, then an occurrence of A in one expression may be replaced with an occurrence of B in another. But such substitution is not valid for propositional attitude verbs. For example, John can know that 17×3 is composite (not prime) but fail to know that 51 is composite, even though in every possible world $17 \times 3 = 51$.

The difficulty is more than a quirk about mathematical expressions. It is widely held that proper names and natural kind terms are rigid designators, that is, that they have the same extension (or none) in every possible world. So substitution of pairs of names (‘Cicero’, ‘Tully’) or natural kind terms (‘water’, ‘ H_2O ’) that have the same extensions should be correct in all contexts. However, John can admire Cicero without admiring Tully (even though Cicero was Tully), and he can know that water is safe to drink, without knowing that H_2O is safe to drink (even though water is H_2O). The distinctions between intensions defined by possible worlds semantics appear to be too crude to explain the failure of substitution in propositional attitude contexts.

5 Alternative approaches

Alternatives to the possible worlds approach in intensional logic have grown out of the desire to resolve problems including that of propositional attitudes. One natural reaction is to dispense with talk of intensions altogether, and to explain John’s different reactions to the ideas that 17×3 and 51 are composite as due to his different relationships with the symbols ‘ 17×3 ’ and ‘51’. However, ‘John knows 17×3 is composite’ cannot report a relation between John and the symbol ‘ 17×3 ’, because this sentence can be true even when John has never seen the symbols ‘ 17×3 ’.

[Max Cresswell](#)’s solution to the problem, following Carnap, is to bring in what he calls structured meanings (1985). The meaning of atomic expressions can be given by a possible worlds account. For example, the meaning of ‘51’ can be the function that picks out 51 in every possible world. For complex expressions, such as ‘ 17×3 ’, the meaning is a structure, $\langle \text{times}, 17, 3 \rangle$, composed of the meanings of the sub-expressions ‘times’, ‘17’ and ‘3’. Note that $\langle \text{times}, 17, 3 \rangle$ is not itself a symbol, but does preserve the structure of ‘ 17×3 ’, thereby allowing the meanings of ‘ 17×3 ’ and ‘51’ to differ.

Since neither ‘Cicero’ nor ‘Tully’ is complex, [Cresswell](#) must use different tactics to manage this case. One strategy he discusses is to analyse ‘John admires Cicero’ as ‘John admires the person he calls Cicero’. (However, since Cresswell believes propositional attitude statements are essentially ambiguous, this is not his final word on the topic.)

Other approaches abandon entirely the possible worlds account of intensions. [Anderson](#), for example, presents an intensional logic whose semantics is based on the notion of synonymy (1984). The idea is that the rules of a language allow us to define the class of synonyms for an atomic expression (that is, the set of expressions with the

same intension), called its ‘concept’. Concepts for complex expressions can be built up from the concepts for simple expressions. Expressions of the language are indexed so that A_1 indicates the concept for expression A . The problem of the failure of substitution of necessary identities in propositional attitude contexts is solved as follows. We assume that the verb ‘admires’ is intensional, so it must take expressions for concepts, not expressions for extensions. So ‘Cicero is Tully’ is represented by $c = t$, while ‘John admires Cicero’ is represented by Ajc_1 , where the index 1 indicates that it is the *concept* of Cicero (c_1) that is at issue. But now Ajt_1 (for ‘John admires Tully’) does not follow, since $c = t$ does not warrant substitution of t_1 for c_1 . On the plausible assumption that ‘Cicero’ and ‘Tully’ are not synonymous (that is, do not have the same intension), the substitution can be blocked as desired. Note, however, that ordinary intuitions rule that ‘water’ is synonymous with ‘ H_2O ’, and that ‘ 17×3 ’ is synonymous with ‘51’. Hence the ordinary notion of synonymy cannot explain the failure of substitution in these cases. For a general solution to the problem, **Anderson** must employ a notion of synonymy that is at odds with ordinary intuitions.

Zalta (1988) lays synonymy aside, and introduces abstract objects to serve as the meanings of intensional expressions. Abstract objects may be thought of as visual images or other mental encodings used to represent features of the world. An extensional term c and its corresponding intensional term c_p belong to the same type, and the subscript p indicates which person does the encoding. Using this machinery, ‘John admires Cicero’ is represented by $A(j, c_j)$, where c_j indicates John’s encoding of Cicero. Note that the encodings named by c_j (‘Cicero’) and t_j (‘Tully’) may be distinct. Since the notation for ‘Cicero = Tully’ ($c = t$) lacks the index j , this claim does not warrant substitution of t_j for c_j in $A(j, c_j)$.

See also: **Intensional entities**; **Logical and mathematical terms**, glossary of

JAMES W. GARSON

References and further reading

- Anderson, C.A.** (1984) ‘General Intensional Logic’, in D. Gabbay and F. Guenther (eds) *Handbook of Philosophical Logic*, Dordrecht: Reidel, vol. 2, 355-85. (A clear and useful introduction to intensional logics, including an account of IL. See §§3, 5.)
- Barwise, J. and Perry, J.** (1983) *Situations and Attitudes*, Cambridge, MA: MIT Press. (An enjoyable introduction to situation semantics, an influential outgrowth of possible worlds semantics for intensional logic.)
- Church, A.** (1951) ‘A Formulation of the Logic of Sense and Denotation’, in *Structure, Method, and Meaning*, New York: Liberal Arts Press. (The seminal work in intensional logic, mentioned in §3.)
- Cresswell, M.J.** (1985) *Structured Meanings*, Cambridge, MA: MIT Press. (Referred to in §5. An accessible account of an intensional logic based on structured meanings.)
- Gallin, D.** (1975) *Intensional and Higher-Order Modal Logic*, Amsterdam: North Holland. (The classic formulation of Montague’s approach to intensional logic. Somewhat technical.)
- Montague, R.** (1974) *Formal Philosophy: Selected Papers of Richard Montague*, ed. R.H. Thomason, New Haven, CT: Yale University Press. (A collection of Richard Montague’s papers prefaced by a very helpful introduction describing the application of intensional logic to natural language.)
- Zalta, E.N.** (1988) *Intensional Logic and the Metaphysics of Intensionality*, Cambridge, MA: MIT Press. (An accessible account of Zalta’s approach to intensional logic described in §5.)

Intensionality

The truth or falsity of many sentences depends only on which things are being talked about. Within intensional contexts, however, truth values also depend on how those things are talked about, not just on which things they are. Philosophers and logicians have offered different analyses of intensional contexts and the behaviour of terms occurring within them.

The extension of a term is the thing or things it picks out: for instance, the extension of ‘the Big Dipper’ is the stellar constellation itself. The intension of a term can be thought of as the way in which it picks out its extension. ‘The Big Dipper’ and ‘the Plough’ have the same extension - a particular constellation - but pick it out in different ways. The two terms have different intensions. There is much debate about how we should actually understand the notion of an intension (see [Intensional entities](#)).

For many classes of sentence, the substitution of one term for another with the same extension leaves the truth value unchanged, irrespective of whether the two terms also have the same intension. This is called intersubstitution *salva veritate* (preserving truth). For instance, if it is true that the Big Dipper consists of seven stars, then it is also true that the Plough consists of seven stars. In a number of important cases, however, terms with the same extension but different intensions are not intersubstitutable *salva veritate*. Consider, for instance:

- 1 Tom believes that the Big Dipper consists of seven stars.
- 2 Tom believes that the Plough consists of seven stars.

It is quite possible that Tom mistakenly believes that ‘the Plough’ refers to a totally different constellation consisting of only six stars, in which case (1) could be true and (2) false. Co-extensive terms cannot be intersubstituted *salva veritate* within the scope of the verb ‘believes’. Such contexts are called intensional.

Other verbs which, like ‘believe’, refer to propositional attitudes - verbs like ‘hope’, ‘desire’, ‘fear’ - also create intensional contexts (see [Propositional attitude statements](#)). Propositional attitudes are *intentional* states (note the ‘t’) which has led some to consider the possible connections between intensionality and intentionality (see [Intentionality §2](#)). The two must be distinguished carefully, however, not least because intensional contexts are also created by non-psychological terms, most importantly modal terms like ‘necessarily’ and ‘possibly’. The intensionality of modal contexts leads to particular difficulties when they also involve quantification (see [Modal logic, philosophical issues in §3](#)); these problems have led to controversies over the interpretation not only of modal terms but also of the quantifiers (see [Modal operators; Quantifiers, substitutional and objectual §1](#)).

The failure of intersubstitutability *salva veritate* raises questions about the behaviour of terms in intensional contexts. Frege (1892) argues that, in intensional contexts, the term ‘the Big Dipper’ in (1) refers not to the Big Dipper as usual but to its own intension (or, in Frege’s terminology, its sense; see [Indirect discourse §1](#); [Proper names §5](#); [Sense and reference §5](#)). Quine (1956, 1961) describes intensional contexts as *referentially opaque*, arguing that terms occurring within them do not refer at all (see [Propositional attitude statements §2](#); [Modal operators §1](#)). Davidson’s (1969) *paratactic* analysis of propositional attitude statements and of indirect discourse, another intensional context, tries to preserve our intuition that terms in intensional contexts work in the same way as in any other contexts (see [Indirect discourse §2](#); [Propositional attitude statements §3](#)).

Many philosophers think that logic should have nothing to do with intensions. But the intensionality of many natural language contexts is hard to deny, and the development of intensional logics suggests that a more tolerant attitude might well pay off (see [Intensional logics](#)).

See also: [Concepts](#); [Reference](#); [Semantics, possible worlds](#)

SIMON CHRISTMAS

References and further reading

Davidson, D. (1969) ‘On Saying That’, *Synthese* 19: 130-46; repr. in D. Davidson and J. Hintikka (eds) *Words and Objections: Essays on the Work of W.V. Quine*, Dordrecht: Reidel, 158-74. (Davidson’s paratactic analysis of indirect discourse.)

- Frege, G.** (1892) 'Über Sinn und Bedeutung', *Zeitschrift für Philosophie und philosophische Kritik* 100: 25-50; trans. M. Black as 'On Sense and Reference', *Translations from the Philosophical Writings of Gottlob Frege*, ed. P.T. Geach and M. Black, Oxford: Blackwell, 1962, 56-78.(Frege's seminal discussion of failures of intersubstitution and distinction between sense and reference.)
- Quine, W.V.** (1956) 'Quantifiers and the Propositional Attitudes', *The Ways of Paradox*, Cambridge MA: Harvard University Press, 1976.(Quine on referential opacity.)
- Quine, W.V.** (1961) 'Reference and Modality', *From a Logical Point of View*, Cambridge, MA: Harvard University Press, 2nd edn, 1980; repr. in L. Linsky (ed.) *Reference and Modality*, Oxford: Oxford University Press, 1971.(Quine on the referential opacity of modal contexts. The Linsky collection contains some fairly difficult discussions.)
- Searle, J.R.** (1983) *Intentionality*, Cambridge: Cambridge University Press.(Chapter 1 contains a useful discussion of the relationship between intenSionality and intenTionality.)

Intuitionism

Ultimately, mathematical intuitionism gets its name and its epistemological parentage from a conviction of Kant: that intuition reveals basic mathematical principles as true *a priori*. Intuitionism's mathematical lineage is that of radical constructivism: constructive in requiring proofs of existential claims to yield provable instances of those claims; radical in seeking a wholesale reconstruction of mathematics. Although partly inspired by Kronecker and Poincaré, twentieth-century intuitionism is dominated by the 'neo-intuitionism' of the Dutch mathematician L.E.J. Brouwer. Brouwer's reworking of analysis, paradigmatic for intuitionism, broke the bounds on traditional constructivism by embracing real numbers given by free choice sequences. Brouwer's theorem - that every real-valued function on a closed, bounded interval is uniformly continuous - brings intuitionism into seeming conflict with results of conventional mathematics.

Despite Brouwer's distaste for logic, formal systems for intuitionism were devised and developments in intuitionistic mathematics began to parallel those in metamathematics. A. Heyting was the first to formalize both intuitionistic logic and arithmetic and to interpret the logic over types of abstract proofs. Tarski, Beth and Kripke each constructed a distinctive class of models for intuitionistic logic. Gödel, in his *Dialectica* interpretation, showed how to view formal intuitionistic arithmetic as a calculus of higher-order functions. S.C. Kleene gave a 'realizability' interpretation to the same theory using codes of recursive functions. In the last decades of the twentieth century, applications of intuitionistic higher-order logic and type theory to category theory and computer science have made these systems objects of intense study. At the same time, philosophers and logicians, under the influence of M. Dummett, have sought to enlist intuitionism under the banner of general antirealist semantics for natural languages.

1 Kantian epistemology, constructivism and 'semi-intuitionism'

The thesis from which mathematical intuitionism draws its name is Kant's: that we exercise a pure intuition to apprehend the correctness of basic mathematical claims directly and *a priori*. Kant defended the thesis - for arithmetic and geometry - in the 'Transcendental Aesthetic' of his first *Critique*. Twentieth-century intuitionists, under the leadership of L.E.J. Brouwer, founder of Dutch or neo-intuitionism, adopted a modification of Kant's view. Intuition in elementary arithmetic and analysis was to be identified with a pure intuition of the subdivision of time intervals. Kant's account of geometry was refused and only elementary arithmetic retained intuitive primacy (see [Mathematics, foundations of](#)). By Brouwer's lights, the role of intuition cannot be understudied by empirical inquiry, linguistic understanding or logical reasoning. Brouwer's intuitionism, then, opposes both empiricism and logicism. Moreover, since intuitionists think to know intuitively principles which are infinitary and contentful, intuitionism is also at odds with finitism and formalism.

Deprecation of logic's influence on mathematics has been characteristic of intuitionism and is reminiscent of the attitudes Schopenhauer expressed in *Die Welt als Wille und Vorstellung* (*The World as Will and Representation*). According to Schopenhauer, mathematics is not best developed *more geometrico*, since deductions from axioms obscure the true grounds for mathematical results. He argued that each mathematical truth *S* has its own intuitable ground of being *G* and that the premier means to intuit the truth of *S* would be to uncover *G* and construct it in time without the aid of explicit deduction. Schopenhauer recommended that mathematics be reformulated accordingly.

Since before the time of Brouwer, intuitionistic mathematics has been constructive. In general, the demand for constructivism is the demand that *E* be respected:

(E) The correctness of existential claim $\exists x A(x)$ is to be guaranteed by warrants from which both an object *a* and a further warrant for $A(a)$ are constructible.

Constructivism is commonly deemed revisionary of conventional mathematics, requiring alteration in standard notions of mathematical objects, canons of adequate proof and fundamental principles of reasoning (see [Constructivism in mathematics](#)). Intuitionists trace their constructive lineage at least as far back as Leopold Kronecker (1823-91), who initiated a programme for arithmetizing higher algebra; in this, he demanded for arithmetic a primacy irreducible to natural science or logic and refused to countenance non-constructive existence proofs.

Writings of the so-called ‘semi-intuitionists’, particularly Poincaré and Borel, exerted a strong influence on Brouwer and his followers. Poincaré, considered the leading European mathematician of his day, also rejected the idea that geometric truths express synthetic judgments known a priori but embraced the intuitive aprioricity of elementary arithmetic, with mathematical induction holding pride of place (see [Poincaré, J.H. §§3, 5](#)). In common with the Dutch intuitionists, Poincaré construed infinite totalities as potential, rather than completed. Accordingly, he rejected, along with Russell’s type theory, Zermelo’s axioms for set theory as guarantors of transfinite arithmetic. Borel was a prominent member of the ‘Paris school’ of mathematicians, which included Baire and Lebesgue. He advocated constructivism to the extent that existence was to be tied to explicit definability: specific functions and real numbers were only to be accepted once they were fully defined. Borel distinguished, as did the young Brouwer, a ‘geometric’ from a ‘practical’ continuum. The former was deemed the object of a general intuition, not analysable into individual reals. The latter was the totality of definable real numbers. It was probably in writings by Borel (or an earlier book by Du Bois Reymond) that Brouwer came upon the idea that real numbers be given by choice sequences of rationals generated randomly, as if by successive throws of a die.

2 Brouwer and Dutch intuitionism

The intuitionism of the twentieth century is that of Brouwer: a radical extension of constructive mathematics incorporating anti-classical principles of logic and analysis. Brouwer held the chair in set theory and axiomatics at the University of Amsterdam from 1912 until 1955. He first achieved fame for discoveries in algebraic topology, particularly proofs of the invariance of dimension and of the fixed-point theorem that bears his name. In articles published between 1907 and 1930, Brouwer combined a battery of influences with an austere personal philosophy of life to create a programme for intuitionistic mathematics. The programme called for trenchant criticism of presuppositions of conventional mathematics as well as thorough reconstruction of arithmetic, analysis and set theory so as to free them of those presuppositions.

In accordance with his subjectivistic, even solipsistic, personal philosophy, Brouwer insisted that intuitionistic mathematics be a private study of mental mathematical constructions. Language and interpersonal communication, indeed all intellectual activity, played diminished roles in Brouwer’s scheme of things. He thought language an instrument so defective as to make mathematicians incapable of conveying the precise sense of their constructions to others. Hence, if mathematics is to retain certainty and exactness, it was not to consist in linguistic representations but in languageless construction. Brouwerian mathematics would not be able to derive its authority from the application to mathematics of a universal logic, valid for all domains of thought.

After Brouwer, the principal exponent of Dutch intuitionism was his student and colleague, Heyting. In addition to major contributions to the formalization of intuitionistic logic, arithmetic and set theory and to the interpretation of that formalism, Heyting pursued studies in intuitionistic algebra, geometry and the theory of Hilbert space. Among early intuitionists were M.J. Belinfante, G.F.C. Griss, H. Freudenthal, B. van Rootselaar and, for a time, H. Weyl. The most distinguished proponents of Dutch intuitionism during the last quarter of the twentieth century are students of Heyting: A.S. Troelstra and D. van Dalen. Notable as well has been a group of intuitionists from Nijmegen including H. de Swart and W. Veldman.

3 Elements of Brouwerian intuitionism

Already in his 1907 dissertation, ‘[On the Foundations of Mathematics](#)’, Brouwer was advancing the intuitionistic idea that mathematics is a ‘free creation’ of the mind unfettered by empirical data or linguistic constraints. Beginning with ‘[The Unreliability of the Logical Principles](#)’ (1908 [1975]), Brouwer took critical aim at the ground rules of conventional mathematics, challenging the validity of the law of the excluded middle (LEM). (The 1908 article may also reflect views held by Brouwer’s teacher and colleague G. Mannoury.)

The validity of LEM does conflict with a natural view of constructivity. At present, it is unknown whether there exist an infinite number of twin primes, that is, pairs of prime numbers that differ by 2. Let the predicate $P(x)$ hold of p just in case $\langle p, p + 2 \rangle$ is the greatest twin prime pair, if such a pair exists, or of 1, if there are an infinite number of twin primes. If LEM is valid, there exists a determinate m such that $P(m)$. However, short of solving the problem of twin primes, no one can provide, in accordance with E, a recipe for computing m in decimal notation. It is noteworthy, first, that Brouwer did not think that this style of reasoning proved the negation of LEM. Rather, he maintained that it affords LEM a ‘weak counterexample’, a proof that the intuitionist will, most likely,

never be able to endorse that law. (Quantified versions of LEM are refuted by intuitionistic continuity principles, described below.) Second, a direct mathematical connection is here forged between the truth of disjunctions and the availability of recipes for deciding an issue.

Brouwer also constructed weak counterexamples to such conventional theorems as that every real is either rational or irrational and that the continuum is totally ordered. In the case of the former, generate in parallel the decimal expansions of π and of ρ , where ρ 's expansion begins 0.33333... but is to be halted when a string of seven 7s appears in the expansion of π . Clearly, ρ cannot be irrational, for then the desired string neither appears nor fails to appear. But, to assert that ρ is rational, E requires that one locate integers m and n such that ρ is their quotient. This is not possible, for we would then know where a string of seven 7s appears in the expansion of π , something now unknown.

Prior to 1920, Brouwer adopted the notion of 'choice sequences', also called 'infinitely proceeding' sequences. These need be neither law-like - governed by computable recipes for generating terms - nor even fully determinate in advance. Brouwer permitted restrictions imposed by a spread law (see below) but, beyond that, nothing about the future course of the sequence may be known, other than the fact that its terms are freely and independently chosen.

Brouwer employed two intuitionistic analogues for the conventional notion of a class of numbers. These are the concepts now called 'species' and 'spread'. Species are properties determined by extensional mathematical predicates and are themselves individuated extensionally: pairs of species are identical when they apply to the same items. Spreads are either species of choice sequences constrained to obey a spread law or species of sequences obtained from these via continuous mappings. A 'spread law' is a decidable property of finite sequences which holds of the empty sequence and culls out from the tree of all sequences a subtree containing branches unbounded in length. A choice sequence is an 'element' of a spread S when it determines a branch in S 's tree. A spread with a tree which always branches finitely often was called by Brouwer a 'fan'.

These notions in hand, Brouwer proved the most celebrated result of intuitionistic mathematics, now known as 'Brouwer's theorem': every real-valued function of a real variable defined on a closed, bounded interval is uniformly continuous. This theorem is known to follow from two distinctively intuitionistic principles. The first, originally isolated by Heyting, is weak continuity (WC-N). Let ' α ' and ' β ' range over choice sequences and ' n ' over natural numbers. Neighbourhoods of α are determined by its initial segments s : β lies in the s -neighbourhood of α ($\beta \in s$) whenever it shares s with α .

(WC-N) If $\forall \alpha \exists n A(\alpha, n)$, then, $\forall \alpha \exists n \exists$ neighbourhood s of α $\forall \beta \in s A(\beta, n)$.

E and the indeterminacy of choice sequences are thought to secure plausibility for WC-N. When we have $\forall \alpha \exists n A(\alpha, n)$, there is to hand a uniform method M for constructing, from each choice sequence α , an n such that $A(\alpha, n)$. Since M is uniform, it can take cognizance, in calculating a specific n from α , of no more than an initial segment of α . After all, we may know nothing more of the course of α than the terms already calculated. Hence, M must yield a proof, for that n , of $\forall \beta \in s A(\beta, n)$. WC-N is alone sufficient to yield strong counterexamples to quantified forms of LEM. Contemporary proofs of Brouwer's theorem also rely on the 'fan theorem', an intuitionistic correlate of König's lemma, which asserts that, with respect to the neighbourhoods just described, every fan is constructively a compact topological space.

In such articles as 'Über Definitionsbereiche von Funktionen' (On Domains of Definitions for Functions, 1927 [1975]), Brouwer attempted a detailed proof of Brouwer's theorem and explained how to deduce the fan theorem from another assertion about spreads, the 'bar theorem'. The latter is, in effect, a principle of 'reverse induction' up from the bottom nodes in a countable, well-founded tree. Let S be a spread. A set of finite sequences B in S is a 'bar' when all α in S have initial segments in B . A set of finite sequences A in S is 'hereditary' if A holds of sequence s in S whenever it holds of all s 's descendants in S . The bar theorem asserts that, whenever A is hereditary and contains a decidable bar B , the empty sequence belongs to A . Unfortunately, Brouwer's attempt to prove it suffered from a non sequitur and is widely considered inconclusive.

Most controversial of Brouwer's innovations, even for fellow intuitionists, was his theory of the 'creative subject'. In articles published after 1947, Brouwer put to work, in defining choice sequences, a representation in temporal

series of the discoveries of an idealized mathematician. From this representation, Brouwer looked to derive strong counterexamples - direct refutations - for findings of conventional analysis, among them that every linear equation with non-zero coefficients has a real solution. Some see the creative subject as a child of Brouwer's subjectivistic philosophy because of Brouwer's assumption that any claim which is permanently unknown to the ideal mathematician is false. Due to difficulties attending its formal analysis, the creative subject remains a subject of debate.

4 Heyting's logic and the proof interpretation

To win a prize offered by the Dutch Mathematical Association, Heyting made the first efforts at a comprehensive formalization for Brouwerian logic, arithmetic and set theory. His system, published in 1930 and 1931, was anticipated in part by the incomplete systems of A.N. Kolmogorov and V.I. Glivenko. Notational details aside, Heyting's predicate logic, now treated as definitive of intuitionistic logic, is equivalent to Gentzen's later natural deduction system NJ. Standard formalizations of intuitionistic predicate logic and elementary arithmetic are still called 'Heyting predicate logic' - or HPL, for short - and 'Heyting arithmetic' - HA - in honour of his achievement.

In a monograph of 1934, Heyting advanced an interpretation of the logic which, he believed, expressed its Brouwerian meaning. He gave, by recursion on claims A , conditions on abstract intuitionistic proofs of A . The 'proof interpretations' for ' $\vee$ ', ' $\rightarrow$ ' and ' $\forall$ ' statements are, roughly, as follows.

- ($\vee$) A proof of $A \vee B$ consists in both a number and a further proof such that, if the number has value 0, the further proof is one of A and, if the number has value 1, the proof is of B .
- ($\rightarrow$) A proof of $A \rightarrow B$ affords a construction converting any proof of A into one of B .
- ($\forall$) A proof of $\forall n A(n)$ affords a construction f such that, for any number m , $f(m)$ proves $A(m)$.

If constructions are computable operations, logics sound for this interpretation may not validate LEM. Let $A(n)$ define an undecidable set of natural numbers. Then, $\forall n (A(n) \vee \neg A(n))$ plausibly lacks proof. Were there such a proof - one satisfying ($\rightarrow$) and ($\forall$) - it would yield, computably in n , an $f(n)$ correctly marking the truth or falsity of $A(n)$, thus contradicting $A(n)$'s undecidability.

The idea of testing logical principles with abstract constructions and proofs is visible in Brouwer's writings. In 1932, Kolmogorov had offered, independently of Heyting, a structurally similar interpretation of intuitionistic reasoning as a 'calculus of problems'. Consequently, the proof interpretation is also known as the 'BHK interpretation', after Brouwer, Heyting and Kolmogorov. The adequacy of certain aspects of the BHK has been a matter of academic dispute. For example, if condition ($\rightarrow$) defines the concept 'abstract proof of $A \rightarrow B$ ', then the definition is impredicative: it contains a quantification - 'any proof of A ' - over all proofs (see [Constructivism in mathematics](#)).

5 Metamathematics: intuitionistic systems and translation theorems

In the first decades after 1945, S.C. Kleene and G. Kreisel, together with their students and colleagues, led a host of metamathematical explorers into traditional intuitionistic territory. Kleene's system for Brouwer's analysis, 'FIM' after Kleene's book with R.E. Vesley, *The Foundations of Intuitionistic Mathematics* (1965), brought to it standard formalization. Among Kreisel's innumerable contributions are formal systems for choice sequences, higher-order arithmetic and the theory of constructions. Of special interest to post-war proof theorists was the system T invented by Gödel to express, for his *Dialectica* interpretation, the behaviours of primitive recursive functionals of finite type (see §6 below).

The 1970s brought out a bumper crop of intuitionistic formal theories for sets, types and higher-order functions. Troelstra (1973a) subjected HAS, a second-order version of Heyting arithmetic containing variables for species and an impredicative comprehension axiom, to logical examination. For computer scientists (see §7), Martin-Löf's unconventional ML type theories acquired a certain celebrity. The ML systems were not invented, as was FIM, to formalize Brouwer's analysis but to represent a mixture of traditional ideas with Martin-Löf's own approach to constructivism. Among the ideas ML assimilates are the BHK interpretation and Frege's distinction between thought and assertion as well as such proof-theoretic mainstays as normalization and W. Howard's 'formulas-as-types'.

As for set theories, IZF - ‘intuitionistic Zermelo-Fraenkel’ - was a formalism of choice for studying universes of Heyting-valued sets (see §6). H. Friedman’s system B was intended to formalize E. Bishop’s constructive analysis but is no stronger, proof-theoretically, than HA. P. Aczel’s ‘constructive Zermelo-Fraenkel’ CZF is translatable into an ML system with a type for well-founded trees. CZF is close kin to J. Myhill’s constructive set theory CST.

In 1933, Gödel - and, independently, Gentzen and Bernays - showed how to embed the theorems of conventional Peano arithmetic (PA) faithfully into the negative ($\neg$, $\exists$ -free) fragment of HA. So, PA is seen to be conservative over HA for negative formulas and, as far as consistency goes, HA is commensurate with its classical cousin. Unknown to Gödel and Gentzen, Kolmogorov had already proposed, in 1925, a similar negative translation for a fragment of HPL. More recently, negative translations have been discovered for higher-order arithmetic, type theory and set theory. Gödel also constructed (in 1933) a translation from Heyting’s propositional logic into the modal calculus S4. H. Rasiowa and R. Sikorski later applied Gödel’s idea to quantified systems. During the 1980s, modal interpretations for intuitionistic arithmetic motivated N. Goodman and S. Shapiro to recommend systems for ‘intensional mathematics’.

6 Metamathematics: algebraic, functional and realizability interpretations

Just as conventional logical signs represent Boolean operations, the logical signs in HPL can also be read algebraically. The relevant algebraic operations act on the open sets of a topological space or, more generally, on elements of Heyting algebras - kinds of distributive lattices with greatest element T . So construed, every formula of HPL gets assigned a topological or lattice-theoretic ‘truth-value’ such that HPL-theorems always evaluate to T . In 1936, S. Jaśkowski seems to have been the first to interpret Heyting’s propositional logic over finite distributive lattices. In the next two years, M. Stone and A. Tarski independently interpreted the logic over topological spaces. Tarski’s article (1938) includes a proof, in conventional metalogic, that every propositional formula which is not an intuitionistic theorem has a countermodel over the open subsets of $\mathbb{R}$. A. Mostowski gave topological semantics to HPL. D.S. Scott (1968) showed how to extend topological interpretations to intuitionistic analysis, thereby constructing models for Brouwer’s theorem. In the 1970s, Scott was among the prime movers in the study of Heyting-valued algebraic semantics for intuitionistic higher-order logic and set theory.

Close relatives of topological models (and of Cohen forcing) are Beth and Kripke models. In 1947, E.W. Beth introduced the idea of interpreting HPL over domains attached to nodes in finitely branching trees. He later argued that Heyting’s logic is sound and complete with respect to his notion of model and that this fact is provable intuitionistically. (In the latter, Beth was shown to have erred.) S. Kripke (1965) proposed, in effect, that formulas be assigned upward closed subsets of partially ordered sets as ‘truth-values’. He gave conventional proofs that HPL is sound and complete with respect to these assignments and that validity in Kripke structures and in Beth structures are equivalent. Both types of models have been extended to other intuitionistic systems, including set theory. Later researchers saw that sets of Kripke or of Beth models can be ‘glued’ together to prove that HA, not to mention other systems, possesses the disjunction property (DP) and explicit definability for numbers (EDN). A system S has DP just in case, whenever A and B are sentences and S derives $A \vee B$, then S derives A or S derives B . S has EDN if, whenever $\exists n.A(n)$ is a sentence that S derives, there is a number term t such that S derives $A(t)$.

The premier functional interpretation for HA was Gödel’s 1959 *Dialectica*, named for the journal in which it first appeared. The point of the interpretation, conceived by Gödel as early as 1941, was to obtain a consistency proof for arithmetic via an extension of Hilbertian finitistic reasoning to higher types. Here, each arithmetic formula is replaced by a quantifier-free translation in the formal system T . The logic of HA is thereby subsumed within combinatorics for higher-type objects. Gödel maintained that the correctness of T is sufficiently evident that the interpretation gives a convincing proof for HA’s consistency. Extensions and applications of Gödel’s ideas were soon forthcoming: in the late 1950s and early 1960s, Kreisel and C. Spector constructed *Dialectica*-style interpretations for analysis; J.-Y. Girard found functional interpretations for HAS in 1971.

Generally, a ‘realizability interpretation’ is an association between formulas A and items i such that i serves, intuitively, to encode information available from a constructive proof of A . When i is so associated with A , it is said to ‘realize’ A or to be a realizability witness for A . On Kleene’s original realizability interpretation from 1945, witnesses are natural numbers that encode the computational information given by A . Witnesses of conditionals, for example, are Turing machine indices such that, if $[[A]]$ and $[[B]]$ are the sets of witnesses for A and B ,

respectively, then $[[A \rightarrow B]]$ is the set of all indices of machines which, given inputs from $[[A]]$, output in $[[B]]$. Kleene proved each HA-theorem to have a non-empty set of witnesses. Realizability also shows HA to be consistent with Church's thesis, the assertion that every arithmetical function is Turing computable. Kleene later devised a realizability interpretation for FIM in which witnesses are functions rather than numbers. He also enlarged the idea of realizability to incorporate his q -realizability and 'slash' interpretations; these permit straightforward proofs of DP and EDN for HA. Many forms of realizability have been invented, extending the concept to most intuitionistic systems.

7 Category theory and computer science

In the 1970s and 1980s, applications of formal systems to category theory and to computer science stimulated new interest in intuitionism. During 1969, F.W. Lawvere isolated axioms for categories which are elementary topoi - generalizations of Grothendieck topoi. Together with M. Tierney, he verified that topoi represent categorial versions of set-theoretic universes which are intuitionistic, rather than conventional, in their logic. The set of truth-values within a topos is always a Heyting, rather than a Boolean, algebra. Research has shown that the theory of interpretations over categories encompasses not only algebraic and topological models but also Kripke models, Beth models and realizability. (See [Category theory, applications to the foundations of mathematics](#).)

Computer scientists have put formal constructive proofs to work extracting programs. As E would have it, a proof p of A is constructive to the extent that there are methods M for extracting from p computational information relevant to A . For instance, if p is a constructive proof of $\forall n \exists m A(n, m)$, M should extract from p a recipe guaranteed to compute a function f such that, for all n , $A(n, f(n))$. The power of high-level programming languages makes it possible to ask for software to play the role of methods M : to extract automatically, from formal proofs p , programs guaranteed to compute functions with the behaviours p specifies. Such software would convert a formal constructive proof that a specification is satisfiable into a program certain to satisfy it. During the 1980s, a number of researchers looked to implement constructive program extraction, among them R.L. Constable and his associates at Cornell (1986), creators of the PRL ('proof refinement logic') systems inspired by Martin-Löf's ML.

8 Intuitionism and antirealism

Since the 1970s, Michael Dummett has been principal sponsor for the view that foundational issues for intuitionism and its logic are properly addressed in philosophical discussions over programmatic 'antirealist' semantics for natural language. Dummett proposed that debates between proponents of conventional and of intuitionistic logics gain a solid footing only when viewed as one skirmish within a wider battle between proponents of realist and of antirealist semantics. Dummett's semantic realist insists that the meaning of statements be explicated by theories associating each statement with bivalent truth-conditions which obtain or fail independently of human knowledge. Dummett's conjectured antirealist would have it that the meanings of statements can only be given in terms of operations people have been taught to employ in verifying or falsifying claims, and not in terms of truth-conditions which may prove ineffable. Intuitionism, then, is supposed to represent an expression of antirealism for which the relevant semantic operations would be those entering into the fashioning of intuitionistic proofs. (See [Antirealism in the philosophy of mathematics §2](#).)

Obstacles to the antirealist programme include the difficulties inherent in designing, even in outline, a coherent antirealist semantics that covers all the territory Dummett has suggested it might cover, including antirealism about psychology and the past. Also, it is in no way clear that features of intuitionistic logic and mathematics can be well understood from the vantage of a semantic theory. It remains to be shown that intuitionistic mathematics, rather than finitism or Markovian constructivism, is the true ally of antirealist semantics. Lastly, it is a concern that Dummett's ideas flout one of the main planks of Brouwer's intuitionism: that mathematics must be independent of metaphysics. For traditional intuitionists, there can be no final explanation for intuitionistic mathematics other than the mathematics itself.

See also [Logical and mathematical terms, glossary of](#)

DAVID CHARLES McCARTY

References and further reading

- Beeson, M.J.** (1985) *Foundations of Constructive Mathematics*, Berlin: Springer.(An overview of the metamathematics of constructive formal systems; requires considerable formal experience.)
- Brouwer, L.E.J.** (1975) *Collected Works*, vol. 1, ed. A. Heyting, Amsterdam: North Holland.(An invaluable compendium of Brouwer's writings, some in English and others in German; the articles 'The Unreliability of the Logical Principles' and 'Über Definitionsbereiche von Funktionen' (On Domains of Definitions for Functions) and sections of his dissertation appear. The level of technical difficulty is high.)
- Brouwer, L.E.J.** (1981) *Brouwer's Cambridge Lectures on Intuitionism*, ed. D. van Dalen, Cambridge: Cambridge University Press.(A compact sample of Brouwer's thinking; well edited but quite technical.)
- Constable, R.L. et al.** (1986) *Implementing Mathematics With the NuPrl Proof Development System*, Englewood Cliffs, NJ: Prentice-Hall.(An admirable introduction to the computer system but no substitute for 'hands-on' experience.)
- Dummett, M.A.E.** (1977) *Elements of Intuitionism*, Oxford: Clarendon Press, 1990.(An influential text in which great emphasis is placed upon the metalogic of intuitionism.)
- Dummett, M.A.E.** (1978) *Truth and Other Enigmas*, London: Duckworth, and Cambridge, MA: Harvard University Press.(Essential reading on intuitionism and antirealism; the preface is especially recommended.)
- Fourman, M.P. and Scott, D.S.** (1979) 'Sheaves and Logic', in M.P. Fourman et al. (eds) *Applications of Sheaves*, Springer Lecture Notes in Mathematics, vol. 753, Berlin: Springer, 302-401.(A mathematically sophisticated but lucid introduction to Heyting-valued interpretations of intuitionistic systems.)
- Gödel, K.** (1986-95) *Collected Works*, ed. S. Feferman et al., New York and Oxford: Oxford University Press, 3 vols.(Authoritative compendium of Gödel's published works and very fine source books. The clear and painstaking introductions are especially recommended.)
- Heijenoort, J. van** (ed.) (1967) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879- 1931*, Cambridge, MA: Harvard University Press.(An invaluable collection of significant articles by Gödel, Hilbert, Brouwer, Skolem and others; technical knowledge often required.)
- Heyting, A.** (1956) *Intuitionism: An Introduction*, Amsterdam: North Holland; 3rd revised edn, 1971. (The first - and still the most charming - popular technical account of Brouwer's intuitionism.)
- Kant, I.** (1781/1787) *Critique of Pure Reason*, trans. N. Kemp Smith, New York: St Martin's Press, 1965.(The 'Transcendental Aesthetic' section includes Kant's highly influential views on a priori intuition.)
- Kleene, S.C.** (1945) 'On the Interpretation of Intuitionistic Number Theory', *Journal of Symbolic Logic* 10: 109-24.(Technical but wonderfully clear and amazingly prescient.)
- Kleene, S.C. and Vesley, R.E.** (1965) *The Foundations of Intuitionistic Mathematics*, Amsterdam: North Holland. (Mathematically exacting, historically important but dauntingly formal.)
- Kreisel, G.** (1965) 'Mathematical Logic', in T.L. Saaty (ed.) *Lectures in Modern Mathematics*, vol. 3, New York: Wiley, 95-195.(A challenging sampler of Kreisel's - and others' - thoughts and theorems on the foundations of mathematics.)
- Kripke, S.A.** (1965) 'Semantical Analyses of Intuitionistic Logic', in J. Crossley and M.A.E. Dummett (eds) *Formal Systems and Recursive Functions*, Amsterdam: North Holland, 92-130.(Includes formal results on the completeness of intuitionistic logic but also nontechnical explications of the model theory.)
- Lambek, J. and Scott, P.** (1986) *Introduction to Higher-Order Categorical Logic*, Cambridge: Cambridge University Press.(A fine introduction to a highly technical subject.)
- Martin-Löf, P.** (1984) *Intuitionistic Type Theory*, Naples: Bibliopolis.(Based on lectures delivered by the author. Offers a helpful introduction to a detailed subject.)
- Rasiowa, H. and Sikorski, R.** (1963) *The Mathematics of Metamathematics*, Warsaw: PWN.(A standard reference on algebraic semantics; requires technical proficiency.)
- Schopenhauer, A.** (1818, 1844) *Die Welt als Wille und Vorstellung*, trans. E.F.J. Payne, *The World as Will and Representation*, New York: Dover, 1969, vol. 1, §15.(Schopenhauer's reflections on mathematics.)
- Scott, D.S.** (1968) 'Extending the Topological Interpretation to Intuitionistic Analysis I', *Compositio Mathematica* 20: 194-210.(A topological treatment of Brouwerian analysis that calls for metamathematical experience.)
- Smorynski, C.** (1973) 'Applications of Kripke Models', in A.S. Troelstra (ed.) *Metamathematical Investigation of Intuitionistic Arithmetic and Analysis*, Lecture Notes in Mathematics, vol. 344, Berlin: Springer, 324-91.(A lucid technical introductory overview to the model theory of intuitionistic arithmetic.)
- Tarski, A.** (1938) 'Der Aussagenkalkül und die Topologie' (The Propositional Calculus and Topology), *Fundamenta Mathematica* 31: 103-34; repr. and trans. in *Logic, Semantics, Metamathematics*, ed. J. Corcoran, Indianapolis, IN: Hackett Publishing Company, 1983, 421-54.(Characteristically careful and clear exposition of early results on topological interpretations.)
- Troelstra, A.S.** (1973a) 'Notes on Intuitionistic Second-Order Arithmetic', in A. Mathias and H. Rogers (eds)

Cambridge Summer School in Mathematical Logic, Berlin: Springer, 171-205.(Remains a fine example of the metamathematics of higher-order intuitionistic systems. A companion piece to Troelstra (1973b).)

Troelstra, A.S. (ed.) (1973b) *Metamathematical Investigation of Intuitionistic Arithmetic and Analysis*, Lecture Notes in Mathematics, vol. 344, Berlin: Springer.(Terse and highly technical. For metamathematicians of the 1970s and 1980s, the bible of intuitionistic systems.)

Troelstra, A.S. and Dalen, D. van (1988) *Constructivism in Mathematics: An Introduction*, Amsterdam: North Holland, 2 vols.(A comprehensive technical overview of the current state of constructive mathematics, with emphasis on intuitionism and mathematical logic.)

Intuitionistic logic and antirealism

The law of excluded middle (LEM) says that every sentence of the form $A \vee \neg A$ (' A or not A ') is logically true. This law is accepted in classical logic, but not in intuitionistic logic. The reason for this difference over logical validity is a deeper difference about truth and meaning. In classical logic, the meanings of the logical connectives are explained by means of the truth tables, and these explanations justify LEM. However, the truth table explanations involve acceptance of the principle of bivalence, that is, the principle that every sentence is either true or false. The intuitionist does not accept bivalence, at least not in mathematics. The reason is the view that mathematical sentences are made true and false by proofs which mathematicians construct. On this view, bivalence can be assumed only if we have a guarantee that for each mathematical sentence, either there is a proof of the truth of the sentence, or a proof of its falsity. But we have no such guarantee. Therefore bivalence is not intuitionistically acceptable, and then neither is LEM.

A realist about mathematics thinks that if a mathematical sentence is true, then it is rendered true by the obtaining of some particular state of affairs, whether or not we can know about it, and if that state of affairs does not obtain, then the sentence is false. The realist further thinks that mathematical reality is fully determinate, in that every mathematical state of affairs determinately either obtains or does not obtain. As a result, the principle of bivalence is taken to hold for mathematical sentences. The intuitionist is usually an antirealist about mathematics, rejecting the idea of a fully determinate, mind-independent mathematical reality.

The intuitionist's view about the truth-conditions of mathematical sentences is not obviously incompatible with realism about mathematical states of affairs. According to Michael Dummett, however, the view about truth-conditions implies antirealism. In Dummett's view, a conflict over realism is fundamentally a conflict about what makes sentences true, and therefore about semantics, for there is no further question about, for example, the existence of a mathematical reality than as a truth ground for mathematical sentences. In this vein Dummett has proposed to take acceptance of bivalence as actually defining a realist position.

If this is right, then both the choice between classical and intuitionistic logic and questions of realism are fundamentally questions of semantics, for whether or not bivalence holds depends on the proper semantics. The question of the proper semantics, in turn, belongs to the theory of meaning. Within the theory of meaning Dummett has laid down general principles, from which he argues that meaning cannot in general consist in bivalent truth-conditions. The principles concern the need for, and the possibility of, manifesting one's knowledge of meaning to other speakers, and the nature of such manifestations. If Dummett's argument is sound, then bivalence cannot be justified directly from semantics, and may not be justifiable at all.

1 Standard intuitionistic logic

The term 'intuitionistic logic' is used in two related senses. In the first sense, intuitionistic logic is the set of logical principles which are valid according to general intuitionism, as originally proposed by the Dutch mathematician L.E.J. Brouwer (1908). In the second, more standard, sense, intuitionistic logic is the logic determined by any one of several equivalent sets of logical rules and axioms. Such an axiom system, motivated by Brouwer's ideas, was first proposed by the Dutch logician Arend Heyting (1930) (see [Intuitionism](#)).

Usually, the phrase 'intuitionistic logic' is taken to mean the logic you get by accepting classical logic and withdrawing your acceptance of the law of excluded middle (LEM). LEM is the law according to which every sentence of the form $A \vee \neg A$ is logically true. Intuitionistically, this law is rejected. This means that the intuitionist is not allowed to assert a sentence $A \vee \neg A$ purely on logical grounds. It does not mean that the negation of that sentence may be asserted. Intuitionistically, as much as classically, the sentence $\neg(A \vee \neg A)$ is a contradiction.

Because of the inconsistency of $\neg(A \vee \neg A)$ we are allowed to infer its negation, $\neg\neg(A \vee \neg A)$. Classically, we could now infer $A \vee \neg A$, because the rule of double negation elimination (DNE), that is, the rule that from $\neg\neg A$ you may infer A , is valid in classical logic. Therefore, DNE is also not intuitionistically valid. By equally simple, intuitionistically acceptable reasoning, you can infer DNE from LEM. In fact, in natural deduction, systems for intuitionistic logic differ from their classical counterparts precisely by not containing DNE.

Classically, LEM is valid in virtue of the truth table explanations of the propositional constants. By the explanation of ‘ $\neg$ ’, either A is true or $\neg A$ is true, and by the explanation of ‘ $\vee$ ’, nothing more is needed for the truth of $A \vee \neg A$. So those who reject LEM must understand these logical constants in a different way.

2 Logic from the intuitionistic viewpoint

L.E.J. Brouwer is commonly regarded as the father of intuitionism, but he himself traces the intuitionist tradition back at least to Kant, who thought of intuitions of space and time as basic to mathematics (see [Kant, I. §5](#)). In Heyting (1934) the following two theses characterize intuitionism:

- (1) Mathematics has not only formal significance, but also content.
- (2) Mathematical objects are immediately grasped by the thinking subject; therefore mathematical knowledge is independent of experience.

Heyting then subdivides intuitionism into ‘half-intuitionism’ and intuitionism proper. The half-intuitionist holds that mathematical objects exist independently of our thinking, but can be known by means of our mathematical constructions. The full-blown intuitionist denies this, or at least rejects the idea that the thesis of mind-independent existence of mathematical objects can be used as a means of mathematical proof.

How can the realistic view of mathematical objects be used as a means of proof? Precisely because that view licenses the use of the law of excluded middle. If mathematical objects exist independently, and have all their properties independently, then, for instance, either, as a matter of fact, the sequence of digits 0123456789 does occur somewhere in the decimal expansion of π , or, as a matter of fact, it does not. For each mathematical statement, either mathematical reality renders that statement true, or else it does not, in which case the statement is false. In this way mathematical realism justifies LEM. LEM, in turn, is used as a means of proof.

Brouwer rejected mathematical realism. He thought that mathematics is an activity where propositions are proved by means of mental constructions, and above all that such mental constructions are precisely what make mathematical propositions true. The statement that 0123456789 does occur somewhere in the decimal expansion of π is rendered true by a construction, for example, by the construction of the decimal expansion up to the point where the sequence occurs, and is rendered false by a construction which proves that it cannot ever occur. And (roughly) a disjunction is rendered true by having one of its disjuncts rendered true. On this view, we are not in advance justified in asserting that the sequence either does or does not occur, for we have no guarantee, before a proof is actually found, that either we can carry out the verifying construction or else we can carry out the falsifying construction.

On these grounds Brouwer rejected LEM. He went further, in that, in sharpest contrast to the formalist school in the philosophy of mathematics (see [Hilbert’s programme and formalism](#)), he refused to accept logic at all as a separate source of knowledge in mathematics. In particular, that a set of mathematical axioms is consistent is not in itself a good reason, according to Brouwer, for accepting them. The relation between logic and mathematics is the opposite. Logical validity depends on insight in the nature of mathematical truth. And this leads to an understanding of the logical operators themselves.

3 Bivalence and the meaning of the logical constants

The truth table explanations give the meaning of the classical propositional constants. For instance, the meaning of ‘ $\rightarrow$ ’ is given by the fact, or stipulation, that $A \rightarrow B$ is false if A is true and B false, and true in the other three cases. This gives the classical meaning of ‘ $\rightarrow$ ’ for the reason that these four truth-value combinations exhaust all possibilities, which, in turn, involves the principle of *bivalence*, the principle that each statement is either true or false.

Under default assumptions, bivalence holds if and only if LEM is valid. The default assumptions are two: first, that $(A \vee B)$ is true if A is true or B is true. Second, that $\neg A$ is true if A is false.

The intuitionist does not accept the principle of bivalence. Again, this means that the intuitionist is not allowed, without further justification, to assert of a sentence A that it is either true or false. It does not mean that the intuitionist may assert that there are sentences which are neither true nor false. In fact, the question whether there are such sentences is a quite separate issue (see [Vagueness §3](#); [Presupposition §2](#); [Many-valued logics](#),

philosophical issues in). However, the intuitionist as well as the classical logician may adopt the view that some sentences are neither true nor false, for example, because of non-referring singular terms. The modified classical logician will then claim that each sentence is either true, or false, or neither true nor false, and the modified intuitionist will not accept this principle.

The intuitionist also accepts the truth tables, but for the intuitionist they do not explain the meaning of the constants. Since the principle of bivalence is not accepted, it cannot be assumed that the entries in the truth tables exhaust all possibilities. The intuitionist may be allowed to assert $A \rightarrow B$, on the basis of a deduction of B from A , without being allowed to assert that either A is false or B is true.

On the intuitionistic conception, a mathematical proposition is true only if there is a proof of it. Moreover, the *meaning* of a mathematical sentence is closely tied to what counts as a proof of it. This is the leading idea in the Brouwer-Heyting or Brouwer-Heyting-Kolmogorov interpretation of the logical constants. Heyting once suggested that a mathematical sentence expresses an expectation of finding a proof of it. The meaning of ' $\vee$ ' is then recursively explained by saying that a sentence $A \vee B$ expresses an expectation which is fulfilled just in case at least one of the expectations expressed by A and by B is fulfilled.

The modern version of this kind of explanation is due primarily to Michael Dummett (1973). Dummett suggests that we can correlate forms of sentence to forms of proof. There is no one form that is common to all proofs of a sentence $A \vee B$, but we can make a distinction between 'canonical' and 'non-canonical' proofs, proposed by both Dummett and the Swedish logician Dag Prawitz (1974), such that a canonical proof of $A \vee B$ is a proof consisting of either a proof of A or a proof of B . Non-canonical proofs are constructions which give us *effective* methods of finding canonical proofs. An effective method is a method which is guaranteed to yield the result after a finite number of elementary steps.

Then a proof of $A \vee B$ either consists of a proof of A or a proof of B , or of an effective method of finding the one or the other. A canonical proof of $A \& B$ is a pair of proofs, of A and of B respectively. A canonical proof of $A \rightarrow B$ is an effective method of transforming any proof of A into a proof of B . A canonical proof of $\neg A$ is an effective method of transforming any proof of A into a proof of ' $\perp$ ' (absurdity - a sentence constant that is false by definition), and therefore a proof of the impossibility of a proof of A . A canonical proof of $\forall x A$ is an effective method for finding, for any term t , a proof of At . A canonical proof of $\exists x A$ is a proof of At , for some term t .

These are the intended intuitionistic interpretations of the logical constants. It follows from these meaning explanations that a non-canonical proof of $A \vee \neg A$ is an effective method for finding either a proof of A or a proof of $\neg A$. If there is such a non-canonical proof of $A \vee \neg A$, then A is called 'effectively decidable' (in Dummett's terminology, which we follow in §6; more technically, the term applies to predicates, or infinite classes of sentences). However, the notion of an absolutely undecidable sentence, that is, of a sentence A such that neither A nor $\neg A$ can be proved at all, does not make sense intuitionistically. A cannot be proved only if it can be proved that A cannot be proved (since truth amounts to provability), but such a proof constitutes a proof of $\neg A$, contradicting the second assumption that $\neg A$ cannot be proved.

A semantic idea that is closely related to the Brouwer-Heyting-Kolmogorov interpretation is the idea that the meaning of the logical constants is given by the rules for them in natural deduction, as first devised by Gerhard Gentzen (1934) and later developed by Prawitz (1965) (see *Logical constants* §4). In particular there is a close analogy between the introduction rules in natural deduction and the forms of canonical proofs. For instance, the introduction rule for 'Čamp;' is that $A \& B$ may be inferred from the joint premises of A and B . From this alternative perspective, introduction rules are valid by definition, and all other rules are to be justified with respect to these.

From both these kinds of meaning-explanation, standard intuitionistic logic is justifiable and LEM is not. But the question is whether further rules than the standard ones are valid. This could be answered in the negative by a proof of semantic completeness, that is, of the statement that all intuitionistically valid principles are derivable from standard intuitionistic logic. Such a proof would have to invoke the intended interpretations rather than some other semantics, and it would have to employ only intuitionistically valid modes of reasoning. At the time of writing, it is not clear that such a proof can be found.

4 Dummett on antirealism and semantics

Brouwer rejected mathematical realism, rejected the law of excluded middle and adopted the idea that mathematical statements are true in virtue of mental constructions. These views naturally go together, but their interrelations are non-trivial.

Antirealism, that is, rejection of realism, comes in many forms. The most direct form is a denial of the existence of entities, for example, the existence of universals. Brouwer denied the existence of extra-mental mathematical objects. If this is denied, then there is a question of evaluating mathematical sentences. Since we want to preserve standard arithmetical judgments, we must understand a sentence such as ' $2 + 2 = 4$ ' so that neither its meaningfulness, nor its truth, depends on the existence of entities of the rejected kind. One possibility is to interpret the sentence in terms of mental constructions, as Brouwer did, but it is not the only one. You can also reinterpret arithmetical sentences, for example, so as to make them true or false by virtue of facts about numerals and their operators. You can even do it in such a way that, if you are a realist about numerals, the principle of bivalence remains valid. So not every form of antirealism will necessitate a change of logic. It will always depend on the way the sentences are reinterpreted.

The form of view suggested, combining a denial of the existence of entities (numbers) with an affirmation of bivalence for (arithmetical) sentences, has been categorized by Dummett as a form of sophisticated realism, rather than a form of antirealism. On Dummett's view, belief in the determinacy of facts is more central to realism than belief in the existence of entities. A mathematical antirealist can accept the existence of numbers and reject the statement that every mathematical state of affairs either obtains or does not obtain. On this view we just cannot take for granted that every mathematical question has an answer in mathematical reality.

These reflections suggest a central place for the principle of bivalence in questions of realism, and that is just what Dummett has proposed. However, there are two more specific features of Dummett's views on bivalence and realism. First, Dummett places bivalence, as a criterion of realism, in an epistemic perspective. In this perspective, realism is in the first place opposed to the knowability principle, that is, the view that if a sentence is true, then it can be known to be true (see [Meaning and verification §5](#)). If you accept bivalence, then, from the epistemic perspective, what is important is that you hold every sentence to be either true or false whether or not it is possible to find out which. And if you reject bivalence, the default reason, from the epistemic perspective, is that on the one hand you accept the knowability principle and, on the other, you do not think there is any guarantee that each sentence can be either known to be true or known to be false.

Second, Dummett regards acceptance of bivalence not just as characteristic of the realist, but as actually defining what it is to be a realist. Similarly, rejection of bivalence by definition makes you an antirealist. Although Dummett recognizes further aspects of realism and antirealism, bivalence is taken as the main dividing line. In this way realism becomes a semantic issue (hence the terms 'semantic realism' and 'semantic antirealism' in discussion of Dummett's views).

There are three interconnected reasons for this change of perspective. First, Dummett holds that traditional metaphysical disputes are framed in pictorial language, with pictures like that of an ethereal realm of abstract entities. These pictures should be replaced by more precisely stated claims. Secondly, the conflict about, for example, the reality of the past, is a conflict about what makes our sentences about the past true, and therefore the question of the reality of the past will involve the question of the proper semantics for sentences about the past. The realist thinks that a sentence can be made true by past facts even if it is impossible for us to know what they are, and the antirealist rejects this as incoherent. Similarly for other disputes over realism. Third, if the conflict is taken as a conflict about semantics, then it can actually be resolved. It will be resolved once it has been shown what a correct meaning theory for the language in question must look like.

5 The relation between realism and bivalence

However, it may seem that Dummett has overstated the role of bivalence in metaphysics. Semantics is what connects bivalence, for some range of sentences, and the determinacy of facts in the corresponding area. What connection there is depends on the kind of semantics, that is, on how the sentences are interpreted. But it also depends on how well the meanings of the sentences are determined.

Suppose that you interpret sentences about the past according to a correspondence model, so that they are made true or false by past facts or events. The sentence ‘George Washington crossed the Potomac’ is then true, if it is true, in virtue of the event of Washington’s crossing the Potomac. Suppose further that all sentences concerned have a well-determined meaning. If a sentence has a well-determined meaning, then it depends only on reality whether or not the sentence is true.

Under these assumptions realism and bivalence are equivalent. If past reality is fully determinate, so that all past states of affairs either obtained or did not obtain, and the meanings of the sentences are such that this and nothing else is needed for making them true or false, then each sentence is true or false. And similarly, if bivalence holds, so does realism.

You may doubt that reality is fully determinate, suspecting that there might be gaps, as it were, and take this as a reason for accepting the knowability principle, thinking that our only guarantee that some state of affairs does or does not obtain is that we can get to know either that it obtains or that it does not. In that case, under the same initial assumptions, bivalence will fail. Conversely, if you accept the knowability principle for another reason, then you have a reason to reject bivalence, since there is no guarantee that each sentence can be known either to be true or to be false. Again under the same assumptions, you must reject realism.

However, if one of these two assumptions is dropped, then the equivalence is lost. Suppose that you give up the second assumption, about the determinacy of meaning. You may think, for example, that the application of a predicate such as ‘...crossed...’ to a pair of objects, such as Washington and the Potomac, is determinately either correct or incorrect only if it is within our present capacity to decide whether it is the one or the other. If we do not have that capacity, then, on this view, the meaning of the predicate is underspecified for these arguments. This view leads to the rejection of bivalence. We cannot assert that each sentence is either true or false, since if we cannot either verify or falsify a particular sentence, then, on this view, its truth-value is left indeterminate. In this case, however, there are no consequences for reality. Reality may be fully determinate. The failure of bivalence depends on meaning only, not on the world.

We get the same result if the first assumption is dropped. Suppose that we adopt a kind of verificationism for sentences about the past. We interpret them so that they are made true or false not by past facts or events, but by present evidence, that is, by such present facts as we take as grounds for asserting those sentences, such as facts about written documents or memories. Then again we have a reason to reject bivalence, for there is no guarantee that we will, for each sentence about the past, be in possession either of truth-making evidence or of falsity-making evidence. And again nothing seems to follow about realism. If sentences about the past are not made true or false by past facts, then failure of bivalence implies nothing about the determinacy of past reality.

The natural reply to the objection is that we have no conception whatsoever about the reality of the past save as that which is represented by our sentences about the past. If these sentences cannot reasonably be given a semantics according to which they are true or false in virtue of past facts, then the notion of an independent past reality must be rejected, for it is only as the truth-maker of our sentences that we have a reason to be realists about it. Moreover, we cannot then use the sentence ‘Past facts are fully determinate’ for stating an independent metaphysical truth, for this sentence must itself be evaluated according to the proposed alternative semantics.

If this reply is correct, then questions of metaphysics are *eo ipso* questions of semantics. And then, as Dummett thinks, metaphysical conflicts are to be settled within the theory of meaning. Indeed, according to Dummett (1991), expressions of metaphysical pictures of reality have no content at all in addition to the semantic views that go with them.

6 Dummett’s antirealist argument

We have seen that the principle of bivalence is of central importance both for realism and for classical logic. Bivalence itself, for any kind of sentence, or for some particular kind, is to be justified by the proper semantics for the kind of sentence in question. And what the proper semantics is, is a question in the general theory of meaning.

Within the theory of meaning Dummett has, notably in ‘What is a Theory of Meaning? (II)’ (1976), proposed general principles, from which an argument is drawn to the conclusion that a semantics of bivalent truth-conditions cannot be correct.

The first step in Dummett's reasoning is the observation that an account of linguistic meaning must be an account of linguistic understanding, that is, of knowledge of meaning. The thesis that meaning consists of truth-conditions is acceptable if and only if it is acceptable that understanding a sentence consists of knowing its truth-conditions, and similarly for other accounts of meaning.

Second, understanding must be publicly manifestable. That is, if I understand a particular sentence, then I must be able to reveal to my interlocutors that I do. This is because language is a means of communication. If there were ingredients in meaning such that knowledge of them could not be made manifest to others, then speakers could not know whether they attached the same meanings to the same expressions, and then communication, as far as those ingredients in meaning are concerned, would be impossible (see [Private states and language](#)).

Thus, if meaning consists of truth-conditions, knowledge of truth-conditions must be manifestable. But how do you manifest your knowledge of the truth-conditions of a sentence? Sometimes you can manifest your understanding by means of a verbal explanation of the meaning of the sentence, but in this way the problem is transferred to the question of your knowledge of the meaning of the sentences you use in your explanation. When you cannot, without circularity, explain further, then you must manifest your understanding in a more primitive way. Dummett says that such a manifestation is a manifestation of a particular capacity, namely, the capacity to get to know that the sentence is true, if it is true, and to get to know that it is false, if it is false. According to Dummett, this is what counts as manifesting knowledge of truth-conditions.

The next step concerns the possibility of such manifestations. If the sentence in question is effectively decidable (see §3 above), then manifesting knowledge of its truth-conditions is possible in principle, but perhaps not in practice. If a sentence is practically decidable, that is, deciding it is within the scope of ordinary human abilities, then I can manifest my knowledge of its truth-conditions by applying a decision method, like making an observation or a small number of calculations, and arrive at a knowledge of its truth-value. If the sentence is effectively decidable, but not in practice, for example, because the number of calculations needed is too large, then I can at least manifest my knowledge of a decision method for it.

The problem appears, according to Dummett, when we come to sentences which are not effectively decidable. A sentence such as 'Two million days before the birth of Benjamin Franklin, an even number of birds landed on Mount St Helen' is not effectively decidable, for we have no method today which is guaranteed to give the right answer. In this case I do not have a capacity to get to know the truth-value, and therefore I cannot manifest my knowledge of its truth-conditions.

The sentence is clearly meaningful, and I clearly understand it. Since knowledge of meaning must be manifestable, and I cannot manifest my knowledge of bivalent truth-conditions in this case, knowledge of its meaning cannot be knowledge of bivalent truth-conditions. Hence, Dummett concludes, meaning does not in general consist of bivalent truth-conditions.

If this argument is sound, then the principle of bivalence cannot be justified in the most natural way, that is, directly from the proper semantics of the sentences it concerns, for that is not a semantics of bivalent truth-conditions. As Dummett points out, this does not in itself invalidate the bivalence principle, since the argument allows that bivalence may be justified in some other way. It seems plausible, however, that bivalence cannot be justified in any way allowed for by Dummett's meaning-theoretical principles.

If bivalence cannot be justified for some particular class of sentences, then neither can semantic realism for that class of sentences. And, since logical principles are perfectly general and topic neutral, if there is some class for which bivalence fails, then, under the default assumptions about the relation between bivalence and LEM (§3), LEM is not a logical law.

This is the destructive aspect of Dummett's enterprise. It has been criticized by a number of philosophers. One objection, due to McDowell (1987), is that the requirement of (full) manifestability is too strong, and another is that Dummett has a rigid and unjustified view of how knowledge of truth-conditions may be manifested.

The constructive aspect is a project of outlining a kind of semantics that satisfies Dummett's own requirements. The model for such a semantics, according to Dummett, is precisely the intuitionistic interpretation of the logical constants. The question is whether this idea can be generalized, in particular from logic and mathematics to

empirical sentences. Such a generalization would amount to some form of verificationism, that is, a view according to which the meaning of a sentence is somehow given by the method, or methods, of verifying the sentence. There are, however, many ways of verifying a verifiable sentence, and only some are directly connected with its meaning. In accordance with the intuitionistic interpretation of the logical constants we should therefore distinguish between canonical and non-canonical, or direct and indirect, verifications of a sentence. Then we can say that the meaning of a declarative sentence is given by what counts as a direct verification of it. On this model, understanding of a sentence would be manifested by exercising an ability to decide, for every proposed piece of evidence, whether or not it constitutes a verification of the sentence. The point is that we can be credited with such an ability even in cases where we in fact do not know how to find a verification. It is enough that we can recognize it if we find it. Therefore the requirement of manifestability is taken to be fulfilled.

However, it has been claimed by Wright (1987) that if full manifestability is required, then problems begin already with sentences not decidable in practice, for example, sentences whose shortest proofs are too long to be surveyable by humans. But, as Wright points out, a finite upper bound on the length of proofs cannot be acknowledged within intuitionism.

Dummett rejects such finite upper bounds, but in fact leaves it open whether intuitionistic logic meets all the meaning-theoretic requirements. He does not, however, see any *principled* reason for doubting it, as he does in the case of classical logic, with its accompanying realism.

See also: [Antirealism in the philosophy of mathematics](#); [Logical and mathematical terms, glossary of](#); [Meaning and truth](#); [Realism and antirealism](#); [Realism in the philosophy of mathematics](#)

PETER PAGIN

References and further reading

- Brouwer, L.E.J.** (1908) ‘De onbetrouwbaarheid der logische principes’, *Tijdschrift voor wijsbegeerte* 2: 152-8; trans. A. Heyting, ‘The Unreliability of the Logical Principles’, in *Collected Works*, vol. 1, *Philosophy and Foundations of Mathematics*, ed. A. Heyting, Amsterdam and Oxford: North Holland, 1975. (Brouwer’s original proposal that intuitionistic logic is the set of logical principles which are valid according to general intuitionism.)
- Brouwer, L.E.J.** (1913) ‘Intuitionisme en formalisme’, *Wiskundig Tijdschrift* 9: 180-211; trans. A. Dresden, ‘Intuitionism and Formalism’, in P. Benacerraf and H. Putnam (eds) *Philosophy of Mathematics: Selected Readings*, Cambridge, MA: Cambridge University Press, 1964. (One of several papers on the nature of intuitionism and logic, originally read in 1912.)
- Dummett, M.** (1973) ‘The Philosophical Basis of Intuitionistic Logic’, in *Truth and Other Enigmas*, London: Duckworth, 1978. (This paper includes the canonical proof kind of interpretation of the logical constants and presents the antirealist argument.)
- Dummett, M.** (1976) ‘What is a Theory of Meaning? (II)’, in G. Evans and J. McDowell (eds) *Truth and Meaning: Essays in Semantics*, Oxford: Clarendon Press. (Dummett’s classic paper on the theory of meaning, including a presentation of the antirealist argument.)
- Dummett, M.** (1977) *Elements of Intuitionism*, Oxford: Clarendon Press. (A thorough introduction to intuitionistic mathematics and logic. Densely written.)
- Dummett, M.** (1991) *The Logical Basis of Metaphysics*, London: Duckworth. (A comprehensive treatment of logic, meaning and metaphysics; partly technical, partly difficult.)
- Gentzen, G.** (1934) ‘Untersuchungen über das logische Schliessen’, *Mathematische Annalen* 39: 176-210; trans. ‘Investigations into Logical Deduction’, in *Collected Papers*, ed. M.E. Szabó, Amsterdam: North Holland, 1969. (Gentzen’s presentation of natural deduction.)
- Heyting, A.** (1930) ‘Die formalen Regeln der intuitionistischen Logik’ (‘The Formal Rules of Intuitionistic Logic’), in *Sitzungsberichte der Preussische Akademie der Wissenschaften*, Berlin. (Heyting’s formalization of intuitionistic propositional logic.)
- Heyting, A.** (1934) *Mathematische Grundlagenforschung. Intuitionismus. Beweistheorie (Foundations of Mathematics. Intuitionism. Proof Theory)*, Berlin: Springer; repr. 1974. (Includes an excellent 25-page introduction to intuitionism, as known at the time.)
- McDowell, J.** (1987) ‘In Defence of Modesty’, in B. Taylor (ed.) *Michael Dummett: Contributions to Philosophy*, Dordrecht: Nijhoff. (A criticism of Dummett’s requirements on meaning theories, including the requirements on manifestability. The book contains other good papers on Dummett, with extensive replies.)

- McGuinness, B. and Oliveri, G.** (eds) (1994) *The Philosophy of Michael Dummett*, Dordrecht: Kluwer. (A good collection of critical comments on Dummett, with extensive replies.)
- Prawitz, D.** (1965) *Natural Deduction: A Proof Theoretic Study*, Stockholm: Almqvist & Wiksell. (A classic study of properties of natural deduction systems for both classical and intuitionistic logic.)
- Prawitz, D.** (1974) 'On the Idea of a General Proof Theory', *Synthese* 27: 63-77. (Introduces the notion of a canonical argument in the context of defining the concept of a valid argument.)
- Tennant, N.** (1987) *Antirealism and Logic: Truth as Eternal*, Oxford: Clarendon Press. (A well-written and useful introduction to, and overview of, the issues, both philosophical and formal, and a defence of antirealism.)
- Wright, C.** (1987) 'Strict Finitism', in *Realism, Meaning and Truth*, Oxford: Blackwell. (This paper concerns among other things whether the strict finitist has not got equally good arguments against the intuitionist as the intuitionist has against classical logic. The book as a whole is to a great extent concerned with Dummett's projects, both the negative and the positive.)

Lambda calculus

The lambda calculus presents a delimited formal setting for expressing and studying properties of mathematical operations and computer programs. Its syntactical unit is the term. Application terms tu represent supplying argument u to routine t . Abstraction terms $\lambda x.t$ (in which λx binds x in t) represent the explicit definition of an operation from a routine t . Term computation is by reduction: $(\lambda x.t)(u)$ reduces to $t[x/u]$, the result of substituting u for x in t ; this corresponds to operating with t on u . Terms are equal either when they differ only in bound variables or when one converts to the other via a finite chain of reductions and counter-reductions. There are two main varieties of lambda calculi: typed and untyped (or type-free). Any untyped term is applicable to any other and equality between such terms is undecidable. In the typed case, terms are indexed by types and application is permitted only when types conform. Here, equality is decidable. A. Church proposed formalisms for both varieties during the 1930s. He and his students S. Kleene and J. Rosser proved the early syntactical metatheorems linking the untyped calculus with computability theory. In 1969, D. Scott gave the first set-theoretic construction of a functional model of the untyped calculus; since then, many other models have been devised. As for the typed calculus, Gödel's Dialectica interpretation inspired many of its applications to proof theory. Also, the idea, due to H. Curry and W. Howard, of using logical formulas as types reveals a close correspondence between typed terms and derivations in intuitionistic formal systems. In the last decades of the twentieth century, the lambda calculus and its models have been put to many uses in computer science, with functional programming and denotational semantics deserving special mention.

1 Abstraction, application and currying

Gottlob Frege's doctrine of concepts and objects (1893) put abstract mathematical functions, including higher-order functions taking functions as inputs or yielding functions as outputs, at the centre of logic (see Frege, G. §3). He insisted that adequate notation for functions must distinguish them from their values. For example, the expression $x^2 + 3x + 7$, for x a natural number, is ambiguous, denoting both the value of the expression at fixed input x and the function that yields $n^2 + 3n + 7$ for every value n of x . Frege realized that, to disambiguate, it suffices to bind x with a prefix operator, such as λx , to yield

$$\lambda x.x^2 + 3x + 7.(1)$$

(Frege's own notation did not employ λ but the Greek symbol for smooth breathing.) These expressions, conveniently displaying both independent variable and a recipe for computation, Frege took to denote functions (in *extension*). Such terms as (1), now called 'abstraction terms', reflect a Fregean notion of function as grasped by abstraction from its values: $0^2 + 3 \cdot 0 + 7$, $1^2 + 3 \cdot 1 + 7$ and so on. Those values are retrievable from (1) by application: the term

$$(\lambda x.x^2 + 3x + 7)(0)$$

denotes the value of the function on the input 0, which we compute by dropping λx and substituting 0 for x in $x^2 + 3x + 7$ to obtain $0^2 + 3 \cdot 0 + 7 = 7$.

In effect, Frege also recognized that lambda notation renders binary or multiary functions as combinations of unary ones, a conversion known as 'currying' after the logician H. Curry. For instance, a lambda term for addition, $\lambda x \lambda y.x + y$, denotes a function taking its arguments one at a time: first x and then y . Its evaluation at 2 yields a unary function

$$(\lambda x \lambda y.x + y)(2) = \lambda y.2 + y,$$

which can be evaluated at 3:

$$(\lambda y.2 + y)(3) = 2 + 3 = 5$$

to yield the desired result.

Alonzo Church was the first to devise formal systems, which he called 'calculi of lambda-conversion' (1941), devoted primarily to codifying such properties of functions. He introduced lambda notation as a variant of the caret ($\wedge$) employed in Whitehead and Russell's *Principia Mathematica* (1910-13) as abstraction operator. Church

intended his calculi to capture an operational conception of function distinct from the extensional notion enshrined in set theory. Set theory conceives of functions as collections of ordered pairs with domain and range so delimited that application of arbitrary function to arbitrary input is prohibited. Hence, set theory does some injustice to operations which apply universally, such as identity, which maps anything to itself, or that operation taking each entity a to the constantly a -valued function. By Church's lights, expressions of the lambda calculus were to denote operations or 'functions in intension': mathematical rules, routines or programs for computing outputs from inputs which may apply universally, even applying to themselves.

2 A formal calculus of terms

There are two general varieties of lambda calculi: 'typed' and 'untyped' (or 'type-free'), both devised by Church. In either case, the basic syntactic category is that of the 'term' (or λ -term). Atomic terms are either variables or constants. Compound terms are of two kinds. There are 'application' terms: if t and u are both terms, so is tu , denoting the input of u to routine t . And, for explicit functions, there are ' λ -abstractions': if t is a term, so is $\lambda x.t$, where x is a variable and λ binds all occurrences of x . In typed systems, each term is also assigned a type and applications tu are well-formed only if the types of t and of u are suitable. (Further discussion of typed systems is postponed to §6). The untyped lambda calculus suffers no such restriction: self-applications such as xx and $(\lambda x.xx)(\lambda x.xx)$ are permitted. Sample untyped terms include $\lambda x.x$, $\lambda x.a$, $\lambda x \lambda y.x$, $\lambda y.y$. The first gives the identity function, the second (with a constant) the constant a -valued function and the third the first projection function, mapping each item a to the constantly a -valued function $\lambda y.a$. The last yields the same function as the first: terms differing only in bound variables, known as ' α -congruent terms', are identified.

Closed abstraction terms without constants are called 'combinators'. Conventionally, combinators $\lambda x.x$, $\lambda x \lambda y.x$ and $\lambda x \lambda y \lambda z.xz(yz)$ are abbreviated I , K and S , respectively, where I stands for 'identity', K for the *Konstanzfunktion* yielding constant functions, and S for a general composition function or *Verschmelzungsfunktion*. This nomenclature derives from M. Schönfinkel's combinatory logic and signals that these lambda terms perform the same operations as their combinatory namesakes (see [Combinatory logic §1](#)).

3 Computations, reductions and normal terms

It is in no way inaccurate to see lambda calculi as rudimentary programming languages, with terms representing computing recipes and their values. As a programming language, unrestricted application in the untyped lambda calculus realizes von Neumann's idea that programs serve also as data (see [Neumann, J. von](#)). There is a notion of lambda calculus computation, which may, like real-life computation, lead to non-terminating 'infinite loops'. As will be apparent (see below) the lambda calculus programming language is rich enough to express all possible algorithms on the natural numbers and is illuminated by respectable concepts of program meaning or denotation.

The notion of 'reduction' (properly, β -reduction) is the principal analogue within the lambda calculus of finitary computation. It affords an analysis of a process of stepwise evaluation obeying a series of atomic, rule-governed steps. Every such step takes its start from the idea of abstraction term exploited above: $\lambda x.t$ names the recipe which begins computation when λx is removed and a name for an input is substituted where x appears free in t . In the lambda calculus, every atomic reduction step begins with a 'redex', an application term of the form $(\lambda x.t)u$ - setting us the project of evaluating recipe $\lambda x.t$ at input u . One replaces the redex with its 'contractum', term $t[x/u]$, the result of substituting u for all free occurrences of x in t (with bound variable changes to avoid clashes). 'Computing' in the lambda calculus is iterated reduction: term v reduces to w if w results from v via a finite number of one-step reductions or replacements of contractum for redex.

Terms are equal when they are interconvertible via computation. If u is derivable from t via a finite sequence of reductions and counter-reductions (perhaps with changes in bound variables) t β -equals (or β -converts to) u ; symbolically, $t = u$. A lambda theory T is a set of formal equations between terms that extends β -equality: whenever t β -equals u , then $t = u$ belongs to T . Intuitively, correct reductions preserve the meanings of the terms involved; hence, β -equal terms share the same meaning. Lambda theories, as sets of equations, respect those meanings. (Lambda models make rigorous sense of these intuitions about meaning. See §5.)

A series of reductions terminates when no further reduction is possible. Irreducible terms are said to be in 'normal form'. It follows that any variable or constant is in normal form, as is any term $\lambda x.t$ in which t itself is normal.

But, just as some computer programs loop infinitely, some untyped terms never reduce to a normal form: $(\lambda x. xxy)(\lambda x. xxy)$ reduces to $((\lambda x. xxy)(\lambda x. xxy))y$ and further reductions always reproduce the original redex. Though some untyped terms lack a normal form, when a term does so reduce, its normal form is unique, at least up to bound variables. This is a corollary of the crucial ‘Church-Rosser theorem’, which states that, if t reduces to u and to v , then there is a term w to which u and v both reduce.

Notions of equality alternative to β -equality are studied, among them $\beta\eta$ -equality. This relation respects the ‘ η -rule’: that $\lambda x. tx$ reduces to t when t does not contain x free.

4 Arithmetic and undecidability

Church and his students S. Kleene and J. Rosser showed, in effect, that, as a programming system, the untyped lambda calculus is comprehensive in that one can define any computable function within it, once natural numbers are represented. Let 0 be $\lambda x \lambda y. y$, 1 be $\lambda x \lambda y. xy$, 2 be $\lambda x \lambda y. x(xy)$, 3 be $\lambda x \lambda y. x(x(xy))$, and so on. These ‘Church numerals’ possess the attractive property, exemplified by 3, that $3ab$ reduces to $a(a(ab))$. Also, all Church numerals are in normal form. Let $[[n]]$ stand for the Church numeral of n . A natural number function f , which may be partial, is λ -defined by t just in case, whenever $f(n)$ exists, $t[[n]]$ reduces to $[[f(n)]]$, and, if $f(n)$ is undefined, then $t[[n]]$ does not reduce to a normal form. A function is λ -definable when it has a λ -defining term.

During 1933, Church speculated that the λ -definable functions coincide with the algorithmically computable functions. This speculation came to be known as ‘Church’s thesis’. If Church’s thesis is true, then every possible recipe for computing with the integers is λ -defined by a term in the untyped lambda calculus and every computation with that recipe appears as a reduction sequence involving that term. Church and Kleene were able to confirm Church’s thesis by proving that every function which is Gödel-Herbrand general recursive is λ -definable. One proof of this requires the construction of fixed-point combinators: terms which, given t , yield u such that $tu \beta$ -equals u . Y , the ‘paradoxical combinator’ discovered by Curry,

$$\lambda x. (\lambda y. x(yy))(\lambda y. x(yy))$$

is one such; applied to any t , Yt β -equals $t(Yt)$.

As befits a programming system, the untyped lambda calculus exhibits undecidability. In 1936, Church used the calculus to prove the undecidability of first-order logic, resolving Hilbert’s *Entscheidungsproblem* (‘decision problem’; see **Hilbert’s programme and formalism**). A general undecidability theorem for the lambda calculus, first proved by D. Scott in 1963 and rediscovered by Curry in 1969, can be stated in terms of recursive separability. Two sets of terms are ‘recursively separable’ if there is a decidable set containing one and disjoint from the other. The ‘Scott-Curry theorem’ asserts that no two non-empty sets of terms closed under β -equality are recursively separable. A corollary is that untyped β -equality and normalizability are both undecidable.

5 Models

Until the late 1960s, syntactic study dominated research on the untyped lambda calculus. Term models were constructed but their elements were syntactic: equivalence classes of terms under such relations as β -equality. An obstacle to building models in set theory, models which capture Church’s vision of a calculus of functions, was self-application. In standard set theory, self-application of functions is not permitted, since no set-theoretic function lies in its own domain.

In 1969, Scott constructed D_∞ , the first recognizably functional model for the untyped case. Scott realized that terms can be interpreted successfully as continuous (rather than arbitrary) functions on complete partial orders. Here, a complete partial order is a partially ordered set with a least element and with least upper bounds for all directed subsets. A continuous function is one whose output values are determinable by finite approximations. D_∞ is a complete partial order enabling self-application, since it is isomorphic to the space of continuous functions from D_∞ into itself. Each element of D_∞ then plays two roles: as element of the model and as function on the model’s domain.

Other models, some simpler in design, were soon forthcoming. G. Plotkin, in 1972, discovered the fundamental idea behind (what is now called) the graph model $P\omega$. The elements of $P\omega$ are subsets of the natural numbers.

$P\omega$ is not isomorphic to its space of continuous functions but, since that space is embedded within $P\omega$, it interprets self-application. A general notion of lambda model, due to A. Meyer (1982) and Scott (1976), which both D_∞ and $P\omega$ exemplify, is tripartite, comprising non-empty domain D , binary application $\circ$ on D and a unary operation Λ . $\circ$ creates on D a model of combinatory logic while Λ serves as a ‘functionalizer’ mapping each d in D to the function taking a into $d \circ a$ (see [Combinatory logic §2](#)). Soundness and completeness theorems confirm the adequacy of this general notion: every lambda model satisfies all basic principles of β -equality and every lambda theory has a lambda model. Intuitively, then, the meanings preserved by reductions are respected in every lambda model.

6 Typed lambda calculi

In the typed lambda calculus, inputs to a function given by a term, as well as its outputs, are restricted to a particular domain or type. For example, to say that term t has type $A \Rightarrow B$ will mean that its inputs are restricted to members of type A and its outputs to B . Hence, if $\mathbb{N}$ is the domain of natural numbers, the successor function on $\mathbb{N}$ is of type $\mathbb{N} \Rightarrow \mathbb{N}$. Typing renders formal two informal ideas. The first is the notion that every well-defined mathematical function bears a specific domain of inputs and range of outputs. The second is that underlying programming languages in which allowable inputs to a program are restricted to members of a particular type, say, natural numbers or integers. With respect to decidability, typing lambda terms makes for notable differences.

Types are defined inductively: there are atomic types and, whenever α and β are types, so is $\alpha \Rightarrow \beta$. One may interpret atomic types to include the natural numbers or the truth-values and compound types $\alpha \Rightarrow \beta$ as containing functions from type α to type β . Types are then used to index terms. Variables and constants, if any, are assumed typed. Application terms must accord with typing: if t is of type $\alpha \Rightarrow \beta$ and u is of type α , then (tu) is of type β . For abstraction terms, when x is of type α and t of type β , $\lambda x.t$ is of type $\alpha \Rightarrow \beta$.

It is straightforward to define natural concepts of reduction, equality and normality for typed terms. There are no self-applications or paradoxical combinators in typed systems. As A. Turing first proved, every typed term reduces to a normal form. Typed reduction also satisfies a Church-Rosser theorem. In fact, a strong normalization theorem obtains: given a typed term, every legitimate way of reducing it eventually ends in the same normal term. It follows that, in contrast with the untyped calculus, typed equality is decidable. Suitable Church numerals are constructible; the natural number functions definable by typed terms are, as discerned independently by H. Schwichtenberg and R. Statman, precisely the extended polynomials. These include integer polynomials such as $\lambda n.3n^2 + 6n + 4$, and functions defined from them using definitions by polynomial cases, for example, $3n^2 + 6n + 4$ if $n = 3$ and n^4 otherwise.

Proof-theoretic applications of typed lambda calculi are prominent. The calculus is extendible to a formalism T for describing the constructive functionals Gödel used in his *Dialectica* interpretation and proof of arithmetic’s consistency (see [Intuitionism §6](#)). Terms in T admit a natural reduction relation which is Church-Rosser and terms are strongly normalizable.

In the typed calculus, if A, B and C are types, then compound types such as $A \Rightarrow (B \Rightarrow A)$ and

$$[A \Rightarrow (B \Rightarrow C)] \Rightarrow [(A \Rightarrow B) \Rightarrow (A \Rightarrow C)]$$

resemble formulas in implicational logic. Curry and R. Feys (1958) noted that these formulas, which are intuitionistically valid, are types suitable for the combinators K and S , respectively. Indeed, the correspondence is general: typed terms accord precisely with correct derivations in the implicational part of intuitionistic propositional logic with application mirroring *modus ponens* and abstraction conditional proof. Moreover, reductive simplifications in logical derivations coincide with reductions in the lambda calculus. This correlation of terms with proofs, known as the ‘Curry-Howard isomorphism’ or ‘formulas as types’, was successfully extended to first-order logic by W. Howard (1980) and to second-order logic by J.-Y. Girard.

7 Computers and the lambda calculus

Research on the lambda calculus is now closely allied with machine computation. For one thing, lambda calculi exhibit, in pure form, issues of decidability, non-termination and typing which face computer theorists. Also, the lambda calculus formalism has inspired a range of functional programming languages, including J. McCarthy’s

LISP and J. Reynold's GEDANKEN. Third, model theory for the lambda calculus gave rise to both operational and denotational semantics for programming languages. A semantics for a programming language is 'operational' when it explains the meaning of a program in terms of its effect upon the state of an abstract machine; it is 'denotational' when the meaning of a program is a mathematical function denoted by that program. P. Landin, in 1965, employed translation into the lambda calculus to give an operational semantics for the language ALGOL. In the early 1970s, Scott and C. Strachey used the ideas behind Scott's lambda calculus models to give denotational semantics for programming languages exhibiting such phenomena as recursion and non-termination.

See also: [Logical and mathematical terms, glossary of](#)

DAVID CHARLES McCARTY

References and further reading

- Barendregt, H.** (1981) *The Lambda Calculus: Its Syntax and Semantics*, Amsterdam: North Holland, 2nd edn, 1984. (A bible of lambda calculus research; authoritative and detailed.)
- Church, A.** (1941) *The Calculi of Lambda Conversion*, Princeton, NJ: Princeton University Press. (The original monograph on the lambda calculus.)
- Curry, H. and Feys, R.** (1958) *Combinatory Logic*, vol. 1, Amsterdam: North Holland. (Once the premier work on the subject; it remains an excellent reference.)
- Frege, G.** (1893) *Grundgesetze der Arithmetik: begriffsschriftlich abgeleitet*, vol. 1, Jena: Pohle; part 1 trans. and ed. M. Furth, *Basic Laws of Arithmetic: An Exposition of the System*, Berkeley, CA: University of California Press, 1964. (Furth's introduction is extremely helpful.)
- Hindley, J. and Seldin, J.** (1986) *Introduction to Combinators and λ -Calculus*, London Mathematical Society Student Texts no. 1, Cambridge: Cambridge University Press. (An excellent brief introduction to the subject. Highly recommended for beginners.)
- Howard, W.** (1980) 'The Formulas-As-Types Notion of Construction', in J. Hindley and J. Seldin (eds) *To H.B. Curry: Essays on Combinatory Logic, Lambda Calculus and Formalism*, New York: Academic Press, 479-90. (The first publication of notes circulated informally in 1969.)
- Kleene, S.** (1981) 'Origins of Recursive Function Theory', *Annals of the History of Computing* 3: 52- 67. (A fascinating eyewitness account of the birth of the lambda calculus and computability theory at Princeton in the early 1930s.)
- Martin-Löf, P.** (1984) *Intuitionistic Type Theory*, Naples: Bibliopolis. (Among the most influential contemporary developments of typed lambda calculus and formulas-as-types.)
- Meyer, A.** (1982) 'What is a Model of the Lambda Calculus?', *Information and Control* 52: 87-122. (A brilliant and highly readable article on a unitary general semantics for the lambda calculus.)
- Rosser, J.B.** (1984) 'Highlights of the History of the Lambda Calculus', *Annals of the History of Computing* 6: 337-49. (Another fine memoir on the early days of lambda calculus research at Princeton and later developments.)
- Scott, D.S.** (1976) 'Data Types as Lattices', *SIAM Journal of Computing* 5: 522-87. (A characteristically charming exposition of results by Scott, Plotkin and others on the graph model.)
- Stenlund, S.** (1972) *Combinators, λ -Terms and Proof Theory*, Dordrecht: Reidel. (Remains one of the best and most lucid introductions to the proof-theoretical applications of the lambda calculus; highly recommended.)
- Stoy, J.** (1977) *Denotational Semantics: The Scott-Strachey Approach to Programming Language Theory*, Cambridge, MA: MIT Press. (The first book-length introduction to Scott-Strachey semantics for programs.)
- Whitehead, A.N. and Russell, B.A.W.** (1910-13) *Principia Mathematica*, Cambridge: Cambridge University Press, 3 vols; 2nd edn, 1925-7; repr. 1994. (The section 'Propositional Functions' in the introduction's chapter 1 describes functional abstraction.)

Linear logic

Linear logic was introduced by Jean-Yves Girard in 1987. Like classical logic it satisfies the law of the excluded middle and the principle of double negation, but, unlike classical logic, it has non-degenerate models. Models of logics are often given only at the level of provability, in that they provide denotations of formulas. However, we are also interested in models which provide denotations of deductions, or proofs. Given such a model two proofs are said to be equivalent if their denotations are equal. A model is said to be 'degenerate' if there are no formulas for which there exist at least two non-equivalent proofs. It is easy to see that models of classical logic are essentially degenerate because any formula is either true or false and so all proofs of a formula are considered equivalent. The intuitionist approach to this problem involves altering the meaning of the logical connectives but linear logic attacks the very connectives themselves, replacing them with more refined ones. Despite this there are simple translations between classical and linear logic.

One can see the need for such a refinement in another way. Both classical and intuitionistic logics could be said to deal with static truths; both validate the rule of modus ponens: if $A \rightarrow B$ and A , then B ; but both also validate the rule if $A \rightarrow B$ and A , then $A \wedge B$. In mathematics this is correct since a proposition, once verified, remains true - it persists. Many situations do not reflect such persistence but rather have an additional notion of causality. An implication $A \rightarrow B$ should reflect that a state B is accessible from a state A and, moreover, that state A is no longer available once the transition has been made. An example of this phenomenon is in chemistry where an implication $A \rightarrow B$ represents a reaction of components A to yield B . Thus if two hydrogen and one oxygen atoms bond to form a water molecule, they are consumed in the process and are no longer part of the current state. Linear logic provides logical connectives to describe such refined interpretations.

1 Syntax

Given these new interpretations of implication and conjunction, they are given new symbols: ' $\multimap$ ' and ' $\otimes$ ', respectively. Thus the chemical reaction mentioned above would be given symbolically as $(H \otimes H \otimes O) \multimap W$.

Despite the appeal of these refined connectives it is clear that our logic is very weak: a formula can be used once only (hence the term 'linear'). To describe situations of static truths, where formulas can be reused (that is, allowed to persist) or ignored, we introduce a new logical connective '!' where $!A$ means that the formula A may be used as many times as needed (including zero). Thus the traditional notion of implication can be simulated by $(!A) \multimap B$. This idea leads to simple translations of both classical and intuitionistic logics into linear logic.

Given two transitions $A \multimap B$ and $A \multimap C$, we clearly have the transition $(A \otimes A) \multimap (B \otimes C)$, but not $A \multimap (B \otimes C)$. However, we shall define a new transition $A \multimap (B \& C)$, which represents the superimposition of the two transitions. It represents a choice of the resulting state B or C . Despite its disjunctive feel the connective ' $\&$ ' is really a conjunction - both states are being offered but only one may be chosen, via one of the axioms $(A \& B) \multimap A$, $(A \& B) \multimap B$. Consequently linear logic has two forms of conjunction, one ($\otimes$) where the initial states are disjoint and one ($\&$) where they are identical. The former is known as a multiplicative connective and the latter as additive. Symmetrically linear logic has two forms of disjunction, the multiplicative form written ' $\star$ ' and the additive form written ' $\oplus$ '.

Linear formulas are built up from a collection of (positive and negative) atomic formulas, written p and $p^\perp$, by means of the multiplicative connectives, ' $\otimes$ ' and ' $\star$ ', additive connectives, ' $\&$ ' and ' $\oplus$ ' and the so-called exponentials, '!' and its dual '?'. The connectives have units (in the sense that in classical logic true (T) is a unit for conjunction ($\wedge$), as $A \wedge T \equiv A$): two multiplicative units, 1 and $\perp$, and two additive units, $\top$ and 0; such that $A \otimes 1 \equiv A$, $A \star \perp \equiv A$, $A \& \top \equiv A$ and $A \oplus 0 \equiv A$. (The logical symbols used here match Girard's rather idiosyncratic selection, apart from ' $\star$ ' which Girard writes as an upside-down ampersand.)

As with classical logic, negation, written $A^\perp$, is a defined notion given inductively over a composite formula. This is given by:

$$\begin{aligned}
 (p^\perp)^\perp &\text{ is } p \\
 (A \otimes B)^\perp &\text{ is } A^\perp \star B^\perp \\
 (A \star B)^\perp &\text{ is } A^\perp \otimes B^\perp \\
 (A \& B)^\perp &\text{ is } A^\perp \oplus B^\perp \\
 (A \oplus B)^\perp &\text{ is } A^\perp \& B^\perp \\
 1^\perp &\text{ is } \perp \\
 \perp^\perp &\text{ is } 1 \\
 \top^\perp &\text{ is } 0 \\
 0^\perp &\text{ is } \top \\
 (!A)^\perp &\text{ is } ?(A^\perp) \\
 (?A)^\perp &\text{ is } !(A^\perp).
 \end{aligned}$$

Also as with classical logic, we are able to define an implication using negation and disjunction: thus $A \multimap B$ is $A^\perp \star B$.

Gentzen's sequent calculus (see [Natural deduction, tableau and sequent systems §2](#)) is a tool for studying various rules of logics. Sequents are written $\Gamma \vdash \Delta$, where $\Gamma (= A_1, \dots, A_n)$ and $\Delta (= B_1, \dots, B_m)$ are finite sets of formulas. The traditional reading of such sequents is that $A_1 \wedge \dots \wedge A_n$ implies $B_1 \vee \dots \vee B_m$. In the linear setting it says that $A_1 \otimes \dots \otimes A_n$ implies $B_1 \star \dots \star B_m$. However, we can employ a shorthand by considering only sequents of the form $\vdash \Delta$, where general sequents of the form $\Gamma \vdash \Delta$ are simulated using $\vdash \Gamma^\perp, \Delta$. The main rules are as follows.

$$\begin{array}{ll}
 \frac{}{\vdash p, p^\perp} & (\text{identity}) \\
 \frac{\vdash \Gamma, A \quad \vdash A^\perp, \Delta}{\vdash \Gamma, \Delta} & (\text{cut}) \\
 \frac{}{\vdash 1} \quad (1) & \frac{\vdash \Gamma}{\vdash \Gamma, \perp} \quad (\perp) \\
 \frac{\vdash \Gamma, A \quad \vdash \Delta, B}{\vdash \Gamma, \Delta, A \otimes B} & (\text{tensor}) \\
 \frac{\vdash \Gamma, A, B}{\vdash \Gamma, A \star B} & (\text{par}) \\
 \frac{\vdash \Gamma, A \quad \vdash \Gamma, B}{\vdash \Gamma, A \& B} & (\text{with}) \\
 \frac{\vdash \Gamma, A}{\vdash \Gamma, A \oplus B} & (\text{sum}_1) \\
 \frac{\vdash \Gamma, B}{\vdash \Gamma, A \oplus B} & (\text{sum}_2)
 \end{array}$$

The difference between the multiplicative and additive connectives can be seen quite clearly by comparing the 'tensor' and 'with' rules. As is the case normally, writing the sequents in a two-sided fashion, and restricting the right-hand side so as to contain at most one formula (for example, $\Gamma \vdash A$), results in the so-called intuitionistic fragment of linear logic. It should be noted that some of the connectives are not available in the intuitionistic fragment. The sequent calculus formulation enjoys the important metatheoretic property of cut elimination (Gentzen's *Hauptsatz*), that is, there is an algorithm which rewrites any proof into one which does not make use of the 'cut' rule.

Girard also showed how one may refine the sequent calculus formulation further by replacing the sequents with graphs. Thus formulas become nodes and applications of the rules become methods for transforming graphs. For

example, the ‘identity’ rule is replaced with two nodes (p and $p^\perp$) and an arc between them, whereas the tensor rule takes the two nodes A and B and places arcs between them and a new node $A \otimes B$. These graphs are known as ‘proof nets’. Several sequent proofs are represented by a single proof net and so one may think of proof nets as revealing the essential inner structure of a linear logic proof.

2 Models

An important feature of linear logic is that its models need not be degenerate, in the sense given earlier. This is in contrast to classical logic. Using the mathematical language of category theory it is relatively easy to describe abstractly which structures are necessary for a model of linear logic: essentially they form particular categories first identified in the 1970s by Barr (see Barr 1991; [Category theory, introduction to](#)). A further problem is to construct concrete models, which are ideally complete (that is, they contain only denotations of provable formulas and valid deductions).

Girard offered two models of linear logic in his seminal 1987 paper. The first, ‘phase spaces’, is a model based on a certain sort of algebraic structure. The second, ‘coherent spaces’, is a graph-theoretic model. A large number of different models have since been identified and again it should be stressed that all of these models are non-degenerate.

An appealing model was given by Blass (1992), based on earlier work on constructive logic by Lorenzen. Here the meaning of a proposition A is given by specifying a game, or dialogue, between two players, one of whom is trying to assert A and the other refute it. Thus a proposition is interpreted as a game and validity by a winning strategy. The connectives can then be understood as operations on games. For example, the connective ‘ $\otimes$ ’ is interpreted as an operation which interleaves two disjoint games or, in other words, plays them side by side. These refined notions of games have been the subject of some intensive work. Despite this effort, and the intuitive appeal of the approach, these models suffer from being incomplete - they validate additional axioms, mostly ‘weakening’: $A \multimap 1$. Some models validate an additional axiom, known as ‘mix’: $(A \otimes B) \multimap (A \star B)$. The status of this axiom (which can also be given as a sequent calculus rule) is uncertain but may be decided by a convincing (game) model. Indeed providing a convincing complete model still remains an open problem.

3 Decidability

One can gain a useful insight into a logic by considering how difficult it is to decide whether a given formula is provable in that logic. Most first-order logics are undecidable but there is considerable variety when considering propositional logics. Both propositional classical logic and propositional intuitionistic logic are decidable albeit with differing complexity ($\mathcal{NP}$ -complete and $\mathcal{PSPACE}$ -complete, respectively; see [Complexity, computational §4](#)). Given that we can encode both these logics into linear logic one might expect linear logic to be decidable. Surprisingly it has been shown that propositional linear logic is undecidable. Indeed even the intuitionistic fragment of linear logic is undecidable. A considerable amount of work exists considering the decidability and complexity of various fragments of linear logic obtained by selecting a subset of the connectives and units.

4 Applications

As linear logic is a refined version of both classical and intuitionistic logic, it is not surprising that it has been applied to many areas of mathematical logic and theoretical computer science. Linear logic’s refined connectives and notion of proof offer a new understanding of classical logic’s connectives and proof deductions. In particular the notion of proof dynamics - the process of cut elimination mentioned in §1 - can be studied with improved accuracy using proof nets. As there are simple translations from classical to linear logic, it is possible to use the non-degenerate models of linear logic to tease out non-degenerate models for classical logic. Many researchers refer to this process as the extraction of the ‘constructive content of classical logic’.

Logic is used by computer scientists in many ways to represent computation. One method is to represent a program as a proof and computation as the process of simplifying the proof or, in other words, bringing it to a normal form (see [Computability theory §3](#)). Another method is to represent a program as a set of rules and a goal, and computation as the process of searching for a proof of the goal given the rules. Given its refined nature, linear logic sheds new light on both these methods.

Importantly, deriving from some of the ideas of the game models described in §2, linear logic appears naturally to encode a process of computation which is inherently parallel in nature. Realizing this process and relating it to existing models of parallel computation is another important open area.

See also: [Logical and mathematical terms, glossary of](#)

G.M. BIERMAN

References and further reading

- Barr, M.** (1991) ‘ $\star$ -Autonomous Categories and Linear Logic’, *Mathematical Structures in Computer Science* 1: 159-78.(A summary of earlier work by Barr which is re-considered in the light of linear logic.)
- Blass, A.** (1992) ‘A Game Semantics for Linear Logic’, *Annals of Pure and Applied Logic* 56: 183-220.(A detailed presentation of the game model discussed in §2.)
- Girard, J.-Y.** (1987) ‘Linear Logic’, *Theoretical Computer Science* 50: 1-101.(The original paper and still essential reading.)
- Girard, J.-Y., Lafont, Y. and Regnier, L.** (eds) (1995) *Advances in Linear Logic*, Cambridge: Cambridge University Press, esp. 1-42, 109-22.(Essentially a collection of research papers but includes an introduction to linear logic by Girard and a survey of decidability results by Lincoln.)
- Hyland, J.M.E.** (1997) ‘Game Semantics’, in A.M. Pitts and P. Dybjer (eds) *Semantics and Logics of Computation*, Cambridge: Cambridge University Press.(A clear tutorial on how to construct particular game models of linear logic.)
- Troelstra, A.S.** (1992) *Lectures on Linear Logic*, CSLI Lecture Notes 29, Stanford, CA: Center for the Study of Language and Information.(A broad but rapid treatment of linear logic.)

Logic, ancient

Western antiquity produced two great bodies of logical theory - those of Aristotle and the Stoics. Both aim to explain what distinguishes good arguments from bad. Both see that the best arguments are valid and that an argument's validity depends on its form. For both, therefore, logic's business is to identify the valid argument forms. Both theories do this by laying down a small number of basic argument forms - Aristotle's 'perfect syllogisms', the Stoics' 'indemonstrables' - and rigorously deriving other valid forms from them. Both theories also try - though in a less systematic manner - to classify the ways in which an argument can go wrong.

Here the similarities between these two logics end. Their most significant differences can be illustrated by comparing basic argument forms from each. The argument 'Every swan is an animal and every animal is moving, so every swan is moving' has the same form as the argument 'Every musician is human and every human is a substance, so every musician is a substance'. The Aristotelian expression of this form is 'A belongs to all B and B belongs to all C, so A belongs to all C'. In this form the letters 'A', 'B' and 'C' stand for any terms whatever, and 'A belongs to all B' replaces 'Every B is an A'. This represents the Aristotelian approach. Compare it with the following. The argument 'If it is day then it is light, it is day, so it is light' has the same form as the argument 'If Dion walks then Dion moves, Dion walks, so Dion moves'. This form is expressed by the Stoics as 'If the first then the second, the first, so the second'. Here the expressions 'the first' and 'the second' stand for any declarative sentences whatever.

In both cases, the validity of the argument form is tantamount to the validity of all arguments having that form (though the Stoics, unlike Aristotle, require that the precise words used in an argument should recur in its form). But the Aristotelian argument form is different in kind from the Stoic one: while it abstracts from terms, the Stoic form abstracts from sentences. Aristotelian logic is a term logic, Stoic logic a sentential one.

1 Aristotle's logic: general

Aristotle was the first great logician (see [Aristotle §§4-5, 7](#)). His logical writings are traditionally grouped together as the *Organon*, comprising *Categories*, *De interpretatione*, *Prior Analytics*, *Posterior Analytics*, *Topics* and *On Sophistical Refutations*.

Categories and *Topics* share a concern with distinguishing different types of predication. In *Categories* beings are classified into substances, quantities, qualities, relatives, places, times, positions, states, actions and affections. Within any one category, beings are related to one another in tree-structures linked by the 'said of' relation; for example, in the category of substance, animal is said of swan and of horse. Between categories there are relations of 'inherence': for example, colour (a quality) inheres in body (a substance). The 'said of' and inherence relations underlie all true predications. Substances are either 'primary' (for example, this horse) or 'secondary' (for example, horse). Among all beings, primary substances are basic: all others are said of or inhere in primary substances, whereas primary substances neither are said of nor inhere in anything. Though *Categories* is concerned mainly with ontological questions, it also states a few principles of inference, including the following: 'When one thing is predicated of another as subject, everything said of the predicate will also be predicated of the subject'.

Topics divides predications into four types: (1) essential and convertible, (2) essential and non-convertible, (3) non-essential and convertible, (4) non-essential and non-convertible. Any predicate thus stands to its subject as (1) a definition (for example, 'biped land-animal' is the definition of man), (2) a 'genus' or 'differentia' (for example, colour is the genus of white, biped is the differentia of man among land-animals), (3) a 'peculiarity' (for example, 'walking in the gymnasium' would be a peculiarity of a particular man if he were the sole walker there) or (4) an 'accident' (for example, evil is an accident of chance). These four types of predication are traditionally known as the 'predicables', and *Topics* formulates a large number of principles (or, commonplaces, 'topoi') concerning them, for example, 'the genus is predicated essentially of whatever the species is predicated of' (IV 2).

The ninth book of *Topics* is known by the separate title *On Sophistical Refutations*. Aristotle there distinguishes fallacies that depend on linguistic factors (such as lexical or syntactic ambiguity) from those that do not.

Fallacies not depending on linguistic factors are seven. The fallacy of 'accident' attributes to a thing what belongs only to an accident of the thing, for example, supposing Coriscus happens to be approaching, it would be wrong to

infer that you do not know Coriscus, given the premise that you do not know who is approaching. The fallacy of ‘*secundum quid*’ argues from a qualified to an unqualified statement, for example, supposing you have a belief about a non-being, it would be wrong to infer that a non-being *is*, given the premise that a non-being is-an-object-of-belief. The fallacy of ‘*ignoratio elenchi*’ argues to the wrong point, for example, by showing that something is double in height when required to show that it is double in width. The fallacy of ‘*petitio principii*’ or begging the question is committed when someone ‘tries to prove by means of itself what is not known by means of itself’ (*Prior Analytics* II 16). The fallacy of ‘affirming the consequent’ is committed when the direction of implication is confused, for example, by confusing the statement that whatever is created has a beginning with the statement that whatever has a beginning is created. The fallacy of ‘false cause’ arises from the inclusion of redundant material in a *reductio* argument, for example, arguing that the diagonal is incommensurable with the side of a square because if it were commensurable, *and if traversing a distance entailed first traversing half the distance*, then we would be committed to the absurd statement that motion is impossible. The fallacy of ‘many questions’ occurs when what in fact are several questions are treated as if they were just one. (See [Fallacies](#).)

De interpretatione contains a semantic theory of categorical statement-forms. *Prior Analytics* contains the theory of the syllogism. *Posterior Analytics* applies the syllogistic to the special case where a syllogism embodies scientific knowledge. It deals with the orderly exposition of a body of scientific knowledge in an axiomatic deductive system: the axioms of such a system will state how things are in themselves and necessarily, and the deductions’ premises will explain the truth of their conclusions.

2 Aristotle’s semantics

Non-modal propositions are of four forms: ‘A belongs to all B’, ‘A belongs to no B’ (these two are ‘universal’), ‘A belongs to some B’ and ‘A does not belong to some B’ (these two are ‘particular’). The letters ‘A’ and ‘B’ stand for terms, that is, nouns or noun phrases. ‘A’ is predicate and ‘B’ subject in these forms. The first and third forms are ‘affirmative’, the second and fourth ‘negative’ (*Prior Analytics* I 2).

Semantics for universal propositions are given in the so-called ‘*dictum de omni et nullo*’: ‘We use the expression “predicated of all” when none of the subject can be taken of which the other term cannot be said, and we use “predicated of none” likewise’ (I 1).

The universal affirmative and particular negative form a pair of ‘contradictories’, as do the universal negative and particular affirmative. The universal affirmative and universal negative are ‘contraries’. Contradictories cannot both be true, nor can they both be false; contraries cannot both be true but may both be false (*De interpretatione* 7).

‘A belongs to no (or some) B’ implies - Aristotle says ‘converts to’ - ‘B belongs to no (or some) A’. ‘A belongs to all B’ converts to ‘B belongs to some A’ (*Prior Analytics* I 2). This can be so only if the universal affirmative has ‘existential import’, that is, implies that something is A and something is B.

Modal propositions are of the forms ‘A belongs necessarily (or contingently) to every B’, ‘A belongs necessarily (or contingently) to no B’, ‘A belongs necessarily (or contingently) to some B’ and ‘A does not belong necessarily (or contingently) to some B’. Aristotle does not state truth-conditions for modal propositions, but he does define the ‘contingent’ (*endechomenon*) as ‘that which is not necessary but, being assumed, results in nothing impossible’. He also distinguishes two senses in which A might be said to belong contingently to all B: either A is contingent for everything to which B belongs or A is contingent for everything for which B is contingent (I 13).

‘A belongs necessarily to no (or some) B’ converts to ‘B belongs necessarily to no (or some) A’. ‘A belongs contingently to some B’ converts to ‘B belongs contingently to some A’; but ‘A belongs contingently to no B’ does not convert (I 3). Contingency modals are subject to the following distinctive laws: ‘A belongs contingently to all B’ is equivalent to ‘A belongs contingently to no B’, and analogously for particulars (I 13).

3 Aristotle’s syllogistic

Aristotle defines a ‘syllogism’ as ‘discourse in which, certain things being proposed, something other than them results of necessity from their being so’ (I 1). Syllogisms have at least two premises. In a two-premised syllogism, one term (the ‘middle’) occurs in both premises. Any two-premised syllogism is in one of three ‘figures’

depending on whether the middle term is (1) once predicate and once subject or (2) twice predicate or (3) twice subject. The conclusion's predicate is the 'major' term, and its subject the 'minor'. The premises are designated as major or minor depending on which of the conclusion's terms occurs in them.

The non-modal syllogistic is based on four 'perfect' or complete syllogisms in the first figure. These (with their medieval names) are 'Barbara' (A belongs to all B and B to all C, so A belongs to all C), 'Celarent' (A belongs to no B and B to all C, so A belongs to no C), 'Darii' (A belongs to all B and B to some C, so A belongs to some C) and 'Ferio' (A belongs to no B and B to some C, so A does not belong to some C).

Modal syllogisms may be pure (with premises and conclusion all containing the same modality) or mixed. Pure modal syllogisms all derive from pure perfect first-figure syllogisms containing either the mode of necessity or that of contingency throughout. Mixed modal syllogisms all derive from mixed perfect first-figure syllogisms whose major premise and conclusion both contain the mode of necessity (or that of contingency), but whose minor premise is non-modal. Thus, Aristotle counts as perfect, not only the four non-modal syllogisms of figure one, but four pure necessity-syllogisms, four pure contingency-syllogisms, four mixed necessity-syllogisms and four mixed contingency-syllogisms - a total of twenty. Not perfect (or even valid) are those first-figure inferences in which the major premise is non-modal while the minor premise and conclusion contain an identical modality.

The perfect syllogisms function as axioms from which other syllogisms are derived (Aristotle says other syllogisms are 'reduced' to the perfect ones). Aristotle's syllogistic is thus the first axiomatized system in history (predating Euclid).

An incomplete syllogism may be perfected in one of three ways. First, by 'direct reduction':

In the second figure, if the negative proposition is necessary, then the conclusion will be necessary.... Let A be possible for no B, and simply belong to all C. Then, since the negative converts, B is possible for no A. But A belongs to all C; so B is not possible for any C.

(I 10)

Here the first premise is converted, thus producing a pair of premises from which the desired conclusion follows by virtue of a 'Celarent' with necessary major and non-modal minor. Second, by 'indirect reduction':

If A belongs to all C and B does not belong to some C, it is necessary that B does not belong to some A. For if B belongs to all A and A to all C, B belongs to all C; but we supposed it did not belong.

(I 6)

Here, by taking the conclusion's contradictory together with the first premise, we get the second premise's contradictory by virtue of 'Barbara'. Third, by 'ecthesis':

when both A and B belong to all C, A will belong necessarily to some B.... If A and B both belong to all C, and we take one of the Cs, say D, A and B will belong to this, so A belongs to some B.

(I 6)

Here the inference's validity is explained by reference to an individual, D, which is supposed to be both an A and a B: there must be such an individual if A and B both belong to all C, but if there is such an individual then A belongs to some B.

In Book II Aristotle works out some metatheorems for his syllogistic, including the following:

- (1) A syllogism with true premises must have a true conclusion.
- (2) Any most general form of syllogism may have (a) all premises true, (b) some premises true and some false, (c) all premises false and the conclusion true.
- (3) The falsity of all (or some) of the premises of a syllogism stated in its most general form does not necessitate the truth of the conclusion.
- (4) No syllogism stated in its most general form commits a logical *petitio principii*, that is, none has the form ' p, q so r ' where p implies r . This result does not hold for all syllogistic forms: 'A belongs to all B, B belongs to all B, so A belongs to all B' commits a logical *petitio*.
- (5) No syllogism by virtue of its most general form commits an epistemological *petitio* in the sense that knowing

or believing premises of that form implies knowing or believing the corresponding conclusion.

4 Theophrastus

Aristotle's successor Theophrastus also wrote on logic. None of his logical writings survives, but we have reports in ancient sources, notably Alexander of Aphrodisias' commentary on Aristotle's *Prior Analytics*. Theophrastus further developed his master's syllogistic in five main ways. First, he pointed out that first-figure syllogisms may have indirect conclusions, for example, from the premises 'A belongs to all B and B to all C' we can deduce 'C belongs to some A'.

Second, in modal semantics Theophrastus rejected Aristotle's definition of the contingent as two-way possibility, and consequently rejected the equivalence between 'A contingently belongs to all B' and 'A contingently belongs to no B'. He also took 'A contingently belongs to no B' to convert to 'B contingently belongs to no A'.

Third, he considered Aristotle's perfect mixed necessity-syllogisms invalid, requiring instead that the modality of the conclusion be no stronger than the weakest modality in the premises.

Fourth, with regard to 'prosleptic' propositions such as 'A belongs to all of that to none of which B belongs' (briefly discussed by Aristotle), Theophrastus considered them to be equivalent to ordinary categoricals.

Fifth, Theophrastus recognized 'wholly hypothetical' syllogisms such as 'If A then B, if B then C, so if A then C', and arranged them in three figures according to the position of their 'middle'. The letters 'A', 'B' and 'C' here stand for sentences rather than terms.

5 Stoic logic and semantics

Among the Stoics and their predecessors, the most famous logicians were Diodorus Cronus, Philo (the Dialectician) and Chrysippus. We have only a few fragments of their work, plus reports in Sextus Empiricus' *Outlines of Pyrrhonism* and *Against the Logicians*, Diogenes Laertius' *Lives of the Philosophers* and other ancient sources.

The Stoics distinguished four causes of argumentative failure: 'incoherence', as in the argument 'If it is day it is light, wheat is being sold in the market, so Dion is walking'; 'redundancy', as in the argument 'If it is day it is light, it is day, Dion is walking, so it is light'; 'formal invalidity', as in 'If it is day it is light, it is not day, so it is not light'; and 'deficiency', as in 'Either wealth is good or bad, it is not good, so it is bad', which is said to be deficient because the first premise omits mention of the possibility that wealth is neither good nor bad. In the first three cases (but not the fourth) the argument as a whole does not instantiate a valid form, though only the first and third are cases of invalidity.

They also studied paradoxes, including the liar paradox ('If you say that you are lying, and speak the truth, then you are lying'), 'the hooded man' (which, like Aristotle's fallacy of accident, concerns an unrecognized friend concealed beneath a hood), the 'sorites' or heap (which asks how many grains it takes to make a heap; see [Vagueness §2](#)) and 'the horned man' ('What you haven't lost you still have, you haven't lost horns, so you still have horns' - a relative of Aristotle's fallacy of many questions).

Stoic semantics and syllogistic is a propositional logic, dealing with the ways in which one proposition may be formed from one or more propositional components. The forms of propositional composition on which the Stoics concentrated are negation, conjunction, disjunction and the conditional. They recognized that a conjunction 'the first *and* the second' is true if and only if both conjuncts are true. A disjunction 'either the first *or* the second' they took in the exclusive sense, requiring that one and only one disjunct be true.

The semantics of conditionals was hotly debated by the Stoics and other contemporary logicians. According to Philo, a conditional proposition ('If the first then the second') is true if and only if it does not have a true antecedent and a false consequent. According to Diodorus Cronus, a conditional is true if and only if it neither is nor ever was possible for the antecedent to be true and the consequent false. A third semantics for conditionals - 'by connection' - proposed that the antecedent be 'incompatible' with the denial of the consequent. A fourth - 'by suggestion' - proposed, somewhat vaguely, that the consequent be 'potentially in' the antecedent. (See [Philo the Dialectician §2](#); [Diodorus Cronus §§3-4](#); [Consequence, conceptions of §2](#).)

6 Stoic syllogistic

There are five 'indemonstrables':

- (1) If the first then the second, the first, so the second.
- (2) If the first then the second, not the second, so not the first.
- (3) Not both the first and the second, the first, so not the second.
- (4) Either the first or the second, the first, so not the second.
- (5) Either the first or the second, but not the second, so the first.

Further inferences were shown valid by the application of one or more rules to the indemonstrables. There were four of these rules, of which only two are now known. The first is equivalent to Aristotle's rule of indirect reduction: if two propositions imply a third, then either of these together with the negation of the third implies the negation of the remaining one. The third rule is the one which underlies Aristotle's procedure of direct reduction: if two propositions imply a third then the first, together with any propositions that imply the second, implies the third.

The third rule is used in proving the inference 'If the first and the second then the third, not the third, but the first, so not the second'. By the second indemonstrable we have 'If the first and the second then the third, not the third, so not both the first and the second'. Applying the third rule to this and the third indemonstrable we get 'If the first and the second then the third, not the third, but the first, so not the second'.

Both the first and third rules are used in the following (modern) proof of the validity of the Stoic inference 'If you know you are dead then you are not dead, if you know you are dead then you are dead, so you do not know you are dead', whose form is 'If the first then not the second, if the first then the second, so not the first'. Applying the first rule to the first indemonstrable we get (a) 'Not the second, the first, so not if the first then the second'. But, again by the first indemonstrable, we have (b) 'If the first then not the second, the first, so not the second'. Applying the third rule to (a) and (b) we get (c) 'If the first then not the second, the first, so not if the first then the second'. And applying the first rule to (c) we get 'If the first then not the second, if the first then the second, so not the first'.

The Stoics accepted the inference 'If the first then if the first then the second, but the first, so the second'. This is said to be validated by two applications of the first indemonstrable. But strictly speaking - and the Stoics were fond of speaking strictly - all that is validated by that procedure is the inference 'If the first then if the first then the second, but the first, *and the first*, so the second'. What is needed to validate the target inference is a 'rule of contraction': if an inference with a repeated premise is valid then so is the same inference without the repetition. This rule is also tacitly used in the above proof of the 'knowing you are dead' syllogism. The rule of contraction, though required by the Stoic system, is nowhere stated in the surviving sources.

The Stoics recognized that if an argument is valid then the conditional whose antecedent is the conjunction of the argument's premises and whose consequent is its conclusion is necessary. Thus, the validity of the first indemonstrable implies the necessity of the conditional 'If "the first and if the first then the second" then the second'.

Some parts of Aristotelian metatheory - namely metatheorems (2c), (3) and (5) in §3 above - do not carry over to the system of indemonstrables. (2c) Given Philonian semantics for the conditional, not both premises of the first indemonstrable can be false: if 'the first' is false then 'If the first then the second' is true. (3) Consequently, given Philonian semantics for the conditional, if both premises of the first indemonstrable are false, the conclusion must be true. (5) The first indemonstrable is epistemically question-begging on the Philonian reading of the conditional: the premise 'If the first then the second' on this reading could be known only if 'the first' were known to be false (which it could not seeing it is known to be true) or 'the second' were known to be true. (But none of this is the case for non-Philonian readings.)

7 Modern interpretations

The 'traditional' logic books of the nineteenth and early twentieth centuries reduced categorical and hypothetical syllogistic to rules of thumb and deprived it of its systematicity and rigour. The categorical syllogism was treated

as an inference, exemplified by ‘All men are mortal, Socrates is a man, so Socrates is mortal’, an example nowhere to be found in *Prior Analytics*. The modal syllogism, if mentioned at all, was dismissed as useless or incoherent.

Łukasiewicz (1951) was the first to interpret Aristotle’s syllogistic as a formalized axiomatic system, rigorous by the highest standards. He returned to Aristotle’s text, and interpreted the syllogism not as an inference but as a conditional statement whose antecedent conjoins two universal or particular (not singular) statements. He was thereby able to treat the syllogistic as an axiomatic system whose theses have the structure of conditional statements with conjunctive antecedents. Because of this structure, the system required an ‘auxiliary theory’ of propositional logic, specifying the logic of conditionals, conjunctions and negation. Practically none of this auxiliary theory is to be found in Aristotle. From Łukasiewicz’s point of view the Stoics therefore marked an advance on Aristotle in articulating a propositional logic.

Corcoran (1972) rejected Łukasiewicz’s interpretation of Aristotelian syllogistic as an axiomatic system whose theses are conditional statement-forms; instead he treated it as a natural-deduction system, thus avoiding the need for an auxiliary propositional logic. An Aristotelian syllogism is for him not just a set of premises and a conclusion but something with a deductive structure. This interpretation enabled him to give modern formal representations not only of Aristotle’s syllogisms but also of Aristotle’s own manner of proving them. He could view Aristotle’s system ‘from the standpoint of modern logic’ without imputing gross oversights to him.

Smiley’s independent work arrived at these same results, but went further than Corcoran in abandoning the aim that every expressible, semantically valid formula should be a thesis of the system. He did this on two grounds. First, Aristotelian syllogistic is not governed by the ‘rule of expansion’ (if certain premises imply a conclusion then any set that includes those premises implies the conclusion). Second, it does not obey the rule of contraction (see §6). The addition of either of these rules to Aristotle’s system, Smiley showed (1973), leads to un-Aristotelian results. The rule of expansion entails that from two mutually contradictory premises every conclusion follows. Thus ‘A belongs to every B, A does not belong to every B, so C belongs to some D’ should be a syllogism; but it is not. Since ‘A belongs to all C, A belongs to all C, so A belongs to some A’ is a syllogism, contraction would entitle us to drop the repetition and say that ‘A belongs to all C’ implies ‘A belongs to some A’. So ‘A belongs to no A’ would imply ‘A does not belong to all C’. Therefore the premises ‘B belongs to no A, B belongs to all A’ - which imply ‘A belongs to no A’ - would also imply ‘A does not belong to all C’. However, ‘B belongs to no A, B belongs to all A, so A does not belong to all C’ is not an Aristotelian syllogism. On Smiley’s interpretation, a syllogism is a deduction in a sense of ‘deduction’ that requires all premises to be used, and which heeds the number of times a premise is used. Smiley constructed a formal system based on such a notion of deduction, proved its completeness, and constructed a decision-procedure for it. These results allowed him to formalize some of Aristotle’s own metatheory, in particular metatheorem (3) (see §3 above).

Thom (1981, 1996) further developed the Smiley interpretation, showing that Aristotle’s syllogistic does not include proper substitutions in theses: thus, while ‘A belongs to all B, B belongs to all C, so A belongs to all C’ is Aristotelian, ‘A belongs to all B, B belongs to all A, so A belongs to A’ is not. This restriction is needed if the interpreter is to construct a formalization to which all of Aristotle’s metatheory applies, since some of that metatheory holds only for theses stated in their most general form (see (2)-(5) in §3 above). He also showed how to develop Aristotle’s sketchy remarks about ‘ecthesis’ into a natural deduction system of singular syllogisms, on which the whole syllogistic (modal as well as non-modal) can be based.

Modern interpretations of modal syllogistic take their impetus from McCall (1963), who restored faith in the rigour of Aristotle’s modal system by formalizing it axiomatically. Johnson (1989) provided a formal semantics for McCall’s formalization. Johnson’s semantics treats ‘A belongs necessarily to every (or no) B’ as true if and only if the Bs are among (or exclude) the *essential* As; ‘A belongs necessarily to some B (or necessarily does not belong to some B)’ he treats as true if and only if some essential B is an essential A (or is an essential non-A). If, as Johnson required, the truth-conditions of Aristotle’s modal sentences make reference to the essential As, then it seems that modal syllogistic is underpinned by some doctrine of essentialism. This idea was developed by Patterson (1995), who explored philosophical links between the semantics of modal sentences and the doctrine of the predicables.

Modern interpretations of Stoic logic spring from the work of Mates (1953), and Kneale and Kneale (1962). The complex relationship between the Stoic and Aristotelian systems has been studied by Frede (1987).

See also: [Logical and mathematical terms, glossary of](#); [Stoicism §11](#)

PAUL THOM

References and further reading

- Alexander of Aphrodisias** (c.200) *In Aristotelis Analyticorum Priorum Librum I Commentarium* (Greek text), ed. M. Wallies, Berlin: Reimer, 1883; *On Aristotle's Prior Analytics 1.1-7*, trans. J. Barnes, S. Bobzien, K. Flannery and K. Ieroddiakonou, London: Duckworth, 1991. (Includes much material on the Stoics and reports of Theophrastus' logical theories as discussed in §4, but not his discussion of modal syllogisms; excellent bibliography.)
- Aristotle** (c.mid 4th century BC) *The Complete Works of Aristotle: The Revised Oxford Translation*, vol. 1, ed. J. Barnes, Princeton, NJ: Princeton University Press, 1984. (See *Categories* for the theory of categories, *Posterior Analytics* for scientific syllogisms, *On Sophistical Refutations* and *Topics* for predicables and fallacies, all discussed in §1; *De interpretatione* for semantics, as discussed in §§1-2; and *Prior Analytics* for syllogistic (Book I) and metatheory (Book II) as discussed in §§2-3.)
- Barnes, J.** (1984) 'Terms and Sentences: Theophrastus on Hypothetical Syllogisms, Dawes Hicks Lecture on Philosophy 1983', *Proceedings of the British Academy* 69: 279-326. (Further non-technical material on Theophrastus.)
- Corcoran, J.** (1972) 'Ancient Logic and its Modern Interpretations', in J. Corcoran (ed.) *Proceedings of the Buffalo Symposium on Modernist Interpretations of Ancient Logic, 21 and 22 April, 1972*, Dordrecht: Reidel. (Partly technical treatment of Aristotelian syllogistic as a natural deduction system.)
- Diogenes Laertius** (early 3rd century) *Lives of the Philosophers*, trans. R.D. Hicks, *Diogenes Laertius Lives of Eminent Philosophers*, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1925, 2 vols; revised H.S. Long, 1972. (Greek text with facing English translation. Volume 2 includes material on the Stoics (book VII) and a life of Chrysippus.)
- Frede, M.** (1987) 'Stoic vs. Aristotelian Syllogistic', in *Essays in Ancient Philosophy*, Oxford: Clarendon Press. (Non-technical discussion.)
- Johnson, F.** (1989) 'Models for Modal Syllogisms', *Notre Dame Journal of Formal Logic* 30 (2): 271- 84. (Technical treatment; referred to in §7.)
- Kneale, W. and Kneale, M.** (1962) *The Development of Logic*, Oxford: Clarendon Press. (Good introduction to the subject, including an extensive bibliography.)
- Lear, J.** (1980) *Aristotle and Logical Theory*, Cambridge: Cambridge University Press. (Further non-technical reading on some of the topics discussed in §3.)
- Lejewski, C.** (1976) 'On Prosleptic Premisses', *Notre Dame Journal of Formal Logic* 17 (1): 1-18. (Technical discussion of prosleptic propositions, mentioned in §4.)
- Lukasiewicz, J.** (1951) *Aristotle's Syllogistic from the Standpoint of Modern Formal Logic*, Oxford: Oxford University Press; 2nd enlarged edn, 1957. (Partly technical treatment; the first to interpret Aristotle's syllogistic as a rigorous, formalized axiomatic system.)
- Mates, B.** (1953) *Stoic Logic*, Berkeley and Los Angeles, CA: University of California Press. (A modern interpretation with a good introduction and bibliography.)
- McCall, S.** (1963) *Aristotle's Modal Syllogisms*, Amsterdam: North Holland. (Axiomatic formalization of Aristotle's modal system; technical.)
- Nortmann, U.** (1994) 'Does Aristotle's Modal Logic Rest on Metaphysical Assumptions?', *Analvomen* 1: 115-25. (An alternative approach to those of Johnson and Patterson; partly technical.)
- Patterson, R.** (1995) *Aristotle's Modal Logic: Essence and Entailment in the Organon*, Cambridge: Cambridge University Press. (Non-technical treatment; referred to in §7.)
- Patzig, G.** (1968) *Aristotle's Theory of the Syllogism: A Logico-Philological Study of Book A of the Prior Analytics*, trans. J. Barnes, Dordrecht: Reidel. (Non-technical discussions of some of the topics mentioned in §3; good bibliography on non-modal syllogistic.)
- Sextus Empiricus** (c.200) *Outlines of Pyrrhonism*, trans. J. Annas and J. Barnes, Cambridge: Cambridge University Press, 1994. (Material on the Stoics, as discussed in §5.)
- Sextus Empiricus** (c.200) *Against the Logicians*, vol. 2, trans. R.G. Bury, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1933-53. (Material on the Stoics, as discussed in §5.)
- Smiley, T.J.** (1973) 'What is a Syllogism?', *Journal of Philosophical Logic* 2 (1): 136-54. (Technical treatment; referred to in §6.)
- Thom, P.** (1981) *The Syllogism*, Munich: Philosophia. (Largely technical treatment on modern interpretations; good bibliography on non-modal syllogistic.)

Thom, P. (1996) *The Logic of Essentialism: An Interpretation of Aristotle's Modal Syllogistic*, Dordrecht: Kluwer.
(Largely technical; bibliography on modal syllogistic.)

Logic in China

Technically, classical China had semantic theory but no logic. Western historians, confusing logic and theory of language, used the term 'logicians' to describe those philosophers whom the Chinese called the 'name school'. The best known of these were Hui Shi (380-305 BC) and Gongsun Lung (b. 380 BC?). This group now also includes the Later Mohists and the term 'distinction school' (translated as 'dialecticians') has become common.

The importance of the more detailed Mohist work came to light in modern times. The Confucian tradition had lost access to it. Rescuing that text rekindled a long-lost interest in Chinese theories of language. The restored Mohist texts give us a general theory of how words work. A term picks out part of reality. Some terms are more general than others; terms like 'dobbins' or 'horse' or 'object' might pick out the same thing. When we use a term to pick something out, we commit ourselves to using the name to pick out similar things and 'stopping' with the dissimilar. Thus, for each term we learn an 'is this' and an 'is not'. 'Is not' generates an opposite for each name and marks the point of distinction or discrimination.

Chinese doctrine portrays disagreements as arising from different ways of making the distinctions that give rise to opposites. The word *bian* (distinction/dispute) thus came to stand for a philosophical dispute. The Mohists argued that, in a 'distinction/dispute', one party will always be right. For any descriptive term, the thing in question will either be an 'is this' or an 'is not'.

Mohists were realistic about descriptions and the world. Real similarities and differences underlie our language. They rejected the claim that words distort reality; to regard all language as 'perverse', they noted, was 'perverse'. The Mohists failed, however, to give a good account of what similarities and differences should count in making a distinction. Mohists also found that combining terms was semantically fickle. In the simplest case, the compound picked out the sum of what the individual terms did. Classical Chinese lacked pluralization so 'cat-dog' works like 'cats and dogs'. Other compound terms (such as 'white horse') worked as they do in English. The confusion led Gongsun Long to argue, on Confucian grounds, that we could say 'white horse is not horse'.

Confucius' linguistics centred on his proposal to 'rectify names'. Confucius used a code with fixed formulations, and therefore tended to treat moral problems as turning on which terms we use in stating them. The abortion dispute illustrates this well. Both sides agree to the rule 'do not kill an innocent person': the dispute becomes one of whether to use the term 'person' or 'foetus'. In contrast, Mohists argued that we should not alter normal term use to get moral results. We simply accept that guiding compounds may not follow normal use. A thief is a person, but killing a thief (executing) is not killing a person (murdering).

These results bolstered Daoist scepticism about words. We never will fashion a 'constant' dao. According to Zhuangzi, even a realistic theory of language (like that of the Mohists) will not give constant guidance. He drew from Hui Shi's approach to language, which emphasized relative terms such as 'large' and 'small'. We may talk of a large horse (relative to other horses) or a large horsefly (relative to other flies), but 'large' itself has no constant standard of comparison. From the premise, 'all such distinctions are relative', Hui Shi fallaciously concluded that 'reality has no distinctions in itself'. Zhuangzi rejected this conclusion and ridiculed Hui Shi's monism. If we say 'everything is one', then our language attempts to 'point to' everything. If it succeeds, then in addition to the 'one-everything' there is the reference to it. That makes two. The whole consisting of everything and saying so then makes three. Referring to that whole makes four, and the fact that we have referred to it makes five, and so on.

Zhuangzi shifted Hui Shi's focus slightly, and concentrated on 'this' and 'that'. These do refer to things, but each use is different. Language, he argued, is not fixed on the world but on our relationships with it. Each existing language (different ways of making guiding distinctions) is equally natural. Human debate is as natural as the chirping of birds. We cannot appeal to nature to settle our disputes about ethics. The standards are not constant; they are historical, variable and diverse in different moral communities.

Distinctions are real, but we can never know if we have found the right ones. Zhuangzi accepts a real world in which language works. Thus, he celebrates the endless possible ways of distinguishing 'this' from 'not-this'. Some alternatives will certainly work better (assuming our present values) than the one we have now. The problem is

that any standard we could use to decide about that would itself be controversial.

The final word came from Xunzi and his student Han Feizi. The former, a Confucian, understood Zhuangzi's arguments to show that the only standard of correct usage must be convention itself. Thus he renewed Confucian tradition and promoted it politically as the only viable and valid conventional system. He advocated government suppression of dissenting voices who 'confuse language' and 'create new terms'. In the end, only the ruler may change language (and then only the 'descriptive' terms). The standards of social assent and dissent come from the Confucian 'sage-kings'. We must adhere to these as the only acceptable ideals; the alternative is anarchy in moral discourse and, consequently, in society.

Han Feizi, seized on Xunzi's attitude toward coercion while discarding the appeal to ancient tradition. Han Feizi had considerable influence on the draconian Qin emperor who ruthlessly carried out his injunction to stamp out philosophical disputes about ethics. This brought the rich tradition of creative philosophy to an abrupt end; religious thought and scholasticism dominated the rest of Chinese intellectual history.

1 Confucius: rectifying names

To understand the development of theory of language, we must place it in the context of Chinese ethical thought. The central focus of ethical dispute was about *dao*, or 'guiding discourse' (see [Confucian philosophy, Chinese §2](#)). Confucius championed the historical 'guiding discourse' of the sage-kings, and purportedly studied it in the ancient documents that formed the curriculum of his school. The *Liji* (*Book of Rites*) is the paradigm for Confucius' conception of ethics.

Confucius' semantic framework was the relation between language and action, not between language and objects or reality. The implicit role of language was prescriptive rather than descriptive. Confucius, oblivious to ethical criticism of his discourse-based guide, addressed mainly problems in practical interpretation. His theory implied a pragmatic relation between language and objects. In order for a discourse to guide us, we must correctly pin names on the world's 'stuff'.

Confucius argued for imitation as the way to achieve this. Social leaders model the proper use of names in publicly practising the 'ritual'; the people learn from these examples and are then able to follow the code as it applies to them. Confucius called this 'rectifying names', and treated it as the key to good government:

Zilu said, 'The ruler of Wei awaits your taking on administration. What would be Master's priority?' The Master replied, 'Certainly - rectifying names!' ... If names are not rectified then language will not flow. If language does not flow, then affairs cannot be completed. If affairs are not completed, ritual and music will not flourish. If ritual and music do not flourish, punishments and penalties will miss their mark. When punishments and penalties miss their mark, people lack the wherewithal to control hand and foot. Hence a gentleman's words must be acceptable to vocalize and his language must be acceptable as action. A gentleman's language lacks anything that misleads - period.

[\(Analects 13:3\)](#)

The strategy of setting examples threatens a regress unless someone in the chain of models knows what example to set in some other way. Stopping the regress pushes Confucianism toward intuitionism. Confucius seemed to regard a mysterious quality, *ren* (humanity), as the key to correct practical interpretation of the *li* (ritual). 'Humanity' is a moral insight that guides the attribution of terms in following circumstances.

One does not rectify by consulting definitions: no Chinese accounts of language generate or point to anything like the concept 'meaning'. The model of the language-world relation is political. Social authorities tag things for ethical purposes. This tagging facilitates guiding discourse. Two theories of Confucian tagging emerge: one determined by traditional training and one by innate moral intuition.

Classical Chinese grammar reflects this model of language-world relations in its topic-comment structure. In the place of sentences, literary Chinese makes 'comments' on contextually indexed 'stuff'. Classical sentences required no subject. Language appeared as far more context dependent than in modern Western thought, with its focus on the subject-predicate sentence and 'complete thought'.

Implicit cognitive theory also mirrored this topic-comment structure. Western philosophy focused on propositional

belief, a mental state that represents a fact, and knowledge. However, ancient Chinese grammar had neither propositional belief nor knowledge structures. Chinese grammar suggests a person tending to deem some real object (*X*) as *P* (the comment). Knowing is correctly deeming or assigning *P* to *X* (or knowing *X*'s being deemed *P*). Literary Chinese used adjectives, one-place verbs and even nouns as two-place predicates in situations where we would use propositional belief structures: one *P*'s some *X*. An alternative structure uses the Daoist concept of *wei* ('deem-act', the *wei* of *wuwei*, 'lack deeming action'). One, with regard to *X*, 'deems-acts' it *P*.

For these reasons (along with the absence of functional inflection, an ideographic conception of writing and the close grammatical similarity of proper and common nouns, adjectives and even verbs), Chinese linguistic theory focused on the question of what term to assign to things rather than on the propositional units so central to Western theory of language and logic (see [Language, philosophy of](#)). The dominant conception was that a word had a scope or range of application, rather than that of referring to individuals or objects. This tendency reflects the fact that Chinese nouns resemble mass-nouns in having cumulative reference, in lacking both grammatical number and articles, and in being associated with various ways of individuating. The use of sortals for individuation became regularized for Chinese nouns at the end of the classical period.

2 Mozi: language utilitarianism

The natural development of this model in Mozi's early work (and the subsequent elaboration in later Mohism) focuses on *bian* (distinction/dispute) (see [Mohist philosophy](#); [Mozi](#)). A term's use involves an 'is this/right' (*shi*) and an 'is not this/wrong' (*fei*). To learn the term is to learn to 'is this/is not this' appropriately with it. Mozi argued that society should use the pre-conventional or natural 'will' toward benefit (and against harm). This contrast guides the assignment of 'is this/is not this' to words used in social discourse. This interpretive proposal flowed imperceptibly into a proposal to order guiding dialogues differently, to change Confucius' traditional 'guiding discourse'.

The clearest example is Mozi's argument about spirits and fate. General utility dictates that social discourse should include the string 'lack spirits' and 'have spirits'. He represents this conclusion as an example of knowing the *dao* of *you-wu* (have-lack) (see [You-wu](#)). That means making a 'is this/is not this' (*li-hai*) distinction for each of these terms using the standard of 'benefit-harm'. We use either 'have' or 'lack' of things when so doing will lead to general utility.

The implication was initially anti-realistic. Mozi advocated three standards of language use. The first recognized the historical, conventional aspect of language. Language should conform to the guidance intentions of the ancient sage-kings. Second, language standards should be applicable by ordinary people using their 'eyes and ears'. Mozi's favourite examples of *fa* (standards) are measurements: a plumb line, a compass, a square and stakes for plotting where the sun rises and sets (see [Fa](#)). Finally, we should use words in ways that maximize general utility. For Mozi, these standards pull in essentially the same direction. He assumed the people's good motivated the sage kings and he repeatedly likened the 'will of nature' to an objective utility measurement.

These standards govern the content and practice of discourse, regulations, injunctions, maxims and slogans. Including any string in the approved discourse was 'making it constant'. The ideal was a discourse *dao* that could consistently (reliably) and correctly (objectively) guide society. Mozi identified that *dao* as the one that resulted in the greatest utility for the country and its people. Thus, description or assignment of names is a handmaiden to ethics.

To count as the constant *dao*, Mozi's 'benefit harm' standard must itself be 'constant'; it should be a reliable, unambiguous and objectively correct, unchanging standard. He argued that in fact it was, since it came from *tian* (nature) rather than from society, convention or contingent history (see [Tian](#)).

Mozi's attack on conventional guiding discourse led [Mencius](#) to defend Confucianism by postulating an innate moral intuition. Mencius argued that language should not manipulate or guide human action. Guidance should only come from the innate patterns or dispositions in the heart-mind; these include an innate ability to 'is this/is not this' (see [Xin \(heart-and-mind\)](#)). The heart-mind selects the appropriate 'is this/is not this' assignment and thus the appropriate action in real contexts. Society should not distort or reshape those natural moral inclinations. Mencius believed that, left to itself, everyone's heart-mind would innately select the 'correct' action for them.

Laozi pointed in a similar direction but undermined Mencius' optimism. Indeed, we should resist the conventional socialization that comes with language. Learning names both constrains natural spontaneity and creates new and disruptive competitive desires. However, Laozi portrayed natural behaviour as being much more 'primitive' than did Mencius. The realm of social concern would extend no further than the local agrarian village. Laozi also emphasized the anti-language aspect of intuitive guidance more than did Mencius. He hinted that Mencius' idea that some particular moral values were intuitive or innate was a result of confusing the unconscious result of learning a guiding language with innate intuition. Learning names involves training in how to make distinctions and how to 'desire' with them. The names, distinctions, desires and actions are linked distortions of natural spontaneity. Laozi's conclusion is his opening line: no guidance in discourse form is constant (see [Daoist philosophy §2](#)).

3 Rediscovery of later Mohist theory

Before discussing the later Mohists, let us glance at the textual explanation of the loss and recovery of the Mohist Canons. The current 'best' textual theory says the Mohists wrote two 'Canons' (I and II). Each consisted of a series of short maxims; the first half of each Canon was written vertically across the top of a standard-sized book of bamboo strips, and the second half on the bottom. The terse analytic theorems were then keyed to another bamboo book containing longer explanations, examples or arguments for them. The second set of bamboo books was indexed by putting the first character of the claim alongside the first character of the explanation.

When scribes later copied each strip of the canons straight through, they turned this clever system into a puzzle. With no understanding of linguistic theory, they treated the whole corpus as a set of consecutive essays. Since classical Chinese had no punctuation or grammatical inflection, this textual disaster obscured the slogans, jumbled the order and shrouded the indexing principle. The scribes absorbed the indexing character as part of the text of the now orphaned explications.

Given the philosophical sophistication and difficulty of the text, the Mohist school's obliteration at the beginning of China's philosophical Dark Age (roughly 200 BC-1000 AD) and the placement of the Canons in the middle of the most vociferous anti-Confucian classical text, the medieval Confucian orthodoxy had little motivation to tackle the puzzle until the late Qing dynasty. Sun Yirang (1848-1908) found the essential clue to unzipping and analysing the content in a phrase in the middle which read, 'read these horizontally'. Other Chinese scholars tried different reconstructions and this work still goes on. Angus Graham's *Later Mohist Logic, Ethics and Science* first delivered a version of the reconstructed text to Western sinologists in 1978 (Graham 1978).

As his title suggests, Graham shared the common view that the subject matter was logic. He assumed Mohists intended the text as a deductively connected set of definitions and propositions. The statements, however, do not resemble definitions. Genus-species form is infrequent and the slogans fit into a theory of use, not of meaning. They are far from forming a deductive set. Still, Graham's assumption guided his reconstruction in ways that made the content accessible to analysis. Many problems and obscurities remain, but Graham's reconstruction reveals a systematic and reasonably coherent theory of language.

The maxims do deal with central philosophical concepts and, like Chinese dictionaries, frequently give lists of substitution characters or a range of examples. Some slogans are metaphors which the explications exploit; others are helpful ways of re-thinking and reflecting on a familiar concept. In addition to theory of language, intelligible sections of the Canon present fragments of epistemological, geometrical, optical and economic theory.

4 Later Mohists: names and distinctions

The School of Names was technically not a school (see [Mohist philosophy §5](#)). Traditional scholars gave the name to individual thinkers who analysed names in conflicting ways; their motivations reflect the differing trends in pre-analytic political thinking. The later Mohists developed Mozi's pragmatic, reality-based approach to naming. Gongsun Long defended the Confucian ideal of an unambiguous guiding scheme of names (the one produced by rectifying names). This principle, he presumed, dictated a goal of one-name-one-thing. Hui Shi raised sceptical problems designed to show that language 'constancy' across all situations was impossible (a Daoist conclusion).

The later Mohists spoke of four 'objects' of *zhi* (knowing): names, 'stuff', union and deeming-action (see [Zhi](#)). Mozi used an example that helps us fix the relevant concept of 'knowing'. Blind people can 'know-to' produce

utterances like ‘black like coal’ or ‘white as snow’ but cannot distinguish things when placed in front of them. They know names, but not ‘stuff’ or union. Thus they cannot use the names to guide their actions.

The Mohists, as we noted above, lacked a doctrine of ‘belief’. ‘Knowing union’ meant competently assigning terms and descriptions to objects. The Mohists accepted that ‘knowing names’ was conventional knowledge. They stressed, however, that we apply conventions to an external reality, known independently of language. The goal of knowledge remained practical guidance, not representation or picturing. Once conventionally attached to some reality, the inherent similarities and differences determined a term’s application. Conventions presuppose a world-guided way to mark distinctions. Mohists portray name-object relations mereologically: a name applies to a scattered reality determined by *tong-yi* (same-different).

The Mohist terms for ‘reference’ - *ju* (pick out) and *qu* (choose) - had a practical tone. A name picks out some ‘stuff’ from the background. Convention determines which similarities and differences mark the boundary between *shi* (‘is this’, what a name picks out) and *fei* (‘is not this’, what it excludes). In using a name, we commit ourselves to go to some real limit and then stop.

The Mohists argued against the one-name-one-reality principle of rectifying names. Names, the Mohists argued, could be very general (like ‘thing’ itself), or based on similarity classes (like ‘horse’) or applied to only one thing (like ‘John’). They saw no objection in principle to having overlapping scopes and even two names for the same thing (like ‘puppy’ and ‘dog’).

Elaboration and defence of this account of *bian* (distinction) led to complications. How should we expand the account to explain what a string of two *ming* (‘names’, characters) picks out? The model used was the string *niu-ma* (ox-horse). The Mohist took this to pick out a compound stuff, the sum of the range of the two component terms (‘draft animals’). They called the unit a *jian* (whole) and its parts *ti* (substantive parts). This analysis of compounds made them analogous to general terms. However, this treatment raised several questions. In what sense is a compound really two things? Could we not view anything as a compound of more basic stuffs? Is there any fundamental kind of ‘substantive part’?

These questions lead to another: are there any basic *tong-yi* (same-different)? The Mohists gave no clear answer. They noticed many senses in which things can be ‘same - different’; some realities might be different only in being called by different names. Being ‘two’ was necessarily differentiated even though called by one name. Realities could also be the same in the sense of being included in some compound object. Conversely, they could be different in not being included in some ‘substantial part’: they could be the same or different in being in the same place or not, and finally they could be same or different in belonging to the same *lei* (kind) (see [Mohist philosophy §5](#)).

However, the Mohists analysed *lei* in a loose way. Having that with which *tong* (same) was the criteria of being *tonglei* (same kind). Not having ‘same’ was the criteria of not being of a ‘kind’. Although they might initially have intended to limit ‘kind’ to natural kinds, the account generalized it to almost any similarity based grouping of stuffs. Thus the Mohists could refer to oxen and horses as the same ‘kind’. The only clear examples of not-‘kind’ are things so unlike they are not comparable. ‘Which is longer, wood or night?’ they suggest, is an unintelligible question because it compares two different ‘kinds’.

The looseness of this account of classifying buttressed the sceptic’s position (see §§8-9) that the world offered no reliable basis for fundamental distinctions. The Mohists seemed vaguely aware of the difficulty and did condemn *kuangju* (‘wild picking out’, where we use irrelevant ‘same’ or ‘different’ to classify or distinguish). They gave no account, however, of what marks a ‘wild’ or irrelevant way of grouping ‘stuff’.

5 Semantic paradoxes and compound terms

The central term of assessment in the Mohist study was *ke* (assertible). They used it in several related ways. An expression may be described as ‘assertible’ of some object. If this object is introduced by another term, then ‘assertible’ became a way of exploring semantic relations between terms (see [Semantics](#)). Mohists asked whether we can sometimes, always or never describe things picked out by term *X* as *Y*.

The analysis, although in terms of assertibility rather than truth, yielded a familiar and important conclusion against certain forms of relativism (see [Relativism](#)). The Mohists argued that in any dispute involving

‘distinctions’, there will be a ‘winner’. If one disputant claimed it was ox and the other that it was not ox, one would be correct. When one disputant claimed the object was ox and the other that it was horse, they did not count that as a distinction dispute. This was merely a formal result, but the Mohists took it as confirming that the world, not conventions, determined the right designation. The winner is the one whose description *dang* (hit on) it.

Another result was a distant cousin of ‘All sentences are false’. Unlike the classic liar paradox (see [Semantic paradoxes and theories of truth](#)), the universal sentence does have a consistent truth value: it is always false. The Mohist conceptual tools, however, lacked two concepts, namely ‘sentence’ and ‘truth’. Instead they construct the following replacement: with regard to language, deeming it exhaustively inadmissible is inadmissible. They explain this result as arising from one’s own language and suggest the proposer try harder to find acceptable words.

This result undermines the anti-language positions, notably those of Mencius and Laozi. It invalidates any claim that language distorts ‘guidance’. Two similar results undermine different formulations of similar points. One version decries distinctions. To make a distinction is to regard something as ‘is not this’, and to say ‘distinctions are wrong’ is to make a distinction. The Mohists say, ‘to *fei* "fei" is perverse’ (that is, to say ‘saying "wrong" is wrong’ is wrong). To recommend making no distinctions is to make a distinction (between making and not-making) and thus the recommender violates his own recommendation. Similarly, the Mohists observe, to teach that teaching lacks use, lacks use.

The analysis of compound terms produced some of the most striking evidence of the different conceptual structure. The Mohists ask what was assertible of the things picked out by compound terms. One standard case was ‘ox-horse’. The Mohists note that ‘not-ox’ is assertible of ‘ox-horse’ on the same grounds that ‘ox’ is. The explanation goes that part of ox-horse is non-ox, so ‘non-ox’ is ‘assertible’. We can understand the idea by reflecting on another example. We may ask someone how many children they have by asking how many ‘boys girls’ they have. Suppose the answer is ‘three’. Now we may ask how many are boys. The answer may come back, ‘none’. This would be a case in which we could say ‘their boy-girls are not boys’.

However, the Mohists seem to have something stronger in mind. Even if the answer was ‘two girls and a boy’, the Mohists would argue that it would be right to say that (some of) their boy-girls were non-boy. Thus the Mohist concludes that, although we cannot say ox is non-ox or horse is non-horse, we can say intelligibly that ox-horse is non-ox-non-horse. The Mohists explain this result as arising from the fact that ox-horse ‘does not interpenetrate’. In any compound that is ox-horse, the minimal parts will be either ox or horse, but not both. They term it a ‘separable’ compound.

The contrast is hard-white. This is also a compound term but, in this case, the components are inseparable. Wherever you go in that which is appropriately called hard-white, you get both. This is the kind of compound term more familiar in Western languages, the intersection compound (see [Language, philosophy of](#)). The scope of the combined term is an intersection of the scopes of the two component terms. Ox-horse, by contrast, is a sum compound. The scope of the combined term is the union of the scopes of the two component terms.

The Mohists do not give us any rule for distinguishing intersection compounds from union compounds beyond the metaphysical feature of ‘penetrating’ or ‘excluding’. They do not explicitly use the language of scope. They also characterize compounds as ‘inseparable’ and ‘separable’. This curious result arises partly because they treat both nouns and adjectives as ‘names’. Both pick out or distinguish one part of reality from the rest. This leaves us with the impression of arbitrariness in how names form compounds, which Mohists explain via metaphysical ‘penetrability’ or its opposite.

6 Phrase matching

Although the Mohists proposed a realist theory of reference, they embedded this in a further theory that we are guided by language. Thus the fourth object is knowing how to ‘deem act’. The Mohists viewed the object of combining names as guiding action. Thus names pick out ‘stuff’, while strings or *ci* (phrases) convey intentions. This led them to analyse mainly ethical compounds. The other study, of ‘ox-horse’ and ‘hard-white’ compounds (see §5), construed them merely as complex names (though it gave no consistent principle for determining their conventional scope). In the long intact section of this text, the Mohists pursued a different analysis, one that can take a superficially logical form:

Premise: *X* is *Y* (white horse is horse)

Conclusion: *KX* is *KY* (ride white horse is ride horse)

They called this ‘matching phrases’, and argued that it was not reliable. Success would be as follows: whenever the version with simple terms was positive (an ‘is this’ phrase), then the parallel with compound terms should be positive (a *ran* (‘so’) phrase). Conversely, a negative base (*X* is not *Y*: a ‘not this’ phrase) should yield a negative result (*KX* is not *KY*: a ‘not so’ phrase).

The Mohists list the different kinds breakdowns:

- Sometimes an ‘is this’ yields a not-‘so’;
- Sometimes a ‘not this’ yields a ‘so’;
- Sometimes a reference is comprehensive and sometimes not;
- Sometimes one reference is ‘is this’ and another one is ‘not this’

The rest of the essay consists of examples that illustrate the respective outcomes.

Graham, drawing on a part-of-speech analysis of ‘is this’ (subject) and ‘so’ (verb), treated this chapter as evidence that the Mohists discovered the subject-predicate sentence. He translated the procedure as ‘matching sentences’ and treated it as a discussion of logical form. The first two models do rely on syntactic complexes (*X Y ke*) which resemble syllogistic premises, but the latter two do not.

Classical Chinese uses no articles and has no ‘is’ verb. Expressions ending with the particle *ye* (assert) mark descriptive uses of referring expressions (noun-phrases). This signals that one is applying a descriptive term to a contextually selected object, not using the term to pick out the object. Translators typically render such structures in English as ‘(*X*) is *Y*’. In Chinese comments, the topic term (the *X*) is optional. The comment functions the same when the string is simply ‘*Y* assert’ and context supplies the topic. The ‘assert’, in other words, does not link two terms, it links a term to an object.

The *shi-ran* (is this-so) analysis given by the Mohists does fit the examples in a way consistent with the topic-comment analysis. If, on the other hand, we focus on the examples as sentences and treat the pattern as a form of inference, then the Mohist analysis will resemble a kind of algebraic logic. However, consistent with a topic-comment analysis, it may be better to regard it as extending the analysis of the conventional semantic effects of combining ‘names’ to form ‘phrases’.

That the analysis rests on conventional semantics of terms rather than logic of sentences sheds a new light on how the examples work. The Mohists do not use the model to correct conventional reasoning errors; rather, they use what we would conventionally say to determine whether a result is a ‘so’ or a not-‘so’. The most thoroughly illustrated breakdowns are those where an ‘is this’ base produces a not-‘so’. The examples are:

- Someone’s parents are people; one’s serving one’s parents is not one’s serving the people;
- One’s younger brother is a handsome man; one’s loving one’s younger brother is not one’s loving a handsome man;
- A carriage is wood; riding a carriage is not riding wood;
- A boat is wood; entering a boat is not entering wood;
- Robbers are people; abounding in robbers is not abounding in people; lacking robbers is not lacking people.

The Mohists expand on the last example in a way that signals both the ethical importance of the analysis and the nature of the alleged breakdown in parallelism:

- Disliking the abundance of robbers is not disliking the abundance of people.
- Desiring to be without robbers is not desiring to be without people.
- Everyone would agree with these so they should not object if we say ‘robbers are people but killing robbers is not killing people’.

We guess, following Graham, that the Mohists are defending their inherited doctrine of universal love by arguing that it is consistent with the (presumed) practice in Mohist communities of executing thieves.

What the denial amounts to is that, even if naming is objectively constant and reliable, the use of names in descriptions of actions or intentions does not reliably take one from an 'is this' to a 'so'. An execution is not murder. Loving a brother is morally required, loving a handsome man is (presumably) shameful. Serving one's parents is one kind of duty and serving the people another. One does not fulfil the latter in merely doing the former.

What emerges is an alternative strategy for dealing with the problem that Confucius addressed via rectifying names (see §1). The Mohists resist the implication that in executing thieves they must deny that thieves are people. They deny instead that executing thieves is murdering people. Rectifying takes place at the 'phrase' or 'so' level rather than at the 'name' level.

The next set of examples illustrate the converse case, those where we start with a 'not this' base and the result is a 'so':

- To read books is not books; to like reading book is to like book.
- Cockfights are not cocks; to like cockfights is to like cocks.
- About to fall in a well is not falling in a well. To stop one about to fall in a well is to stop one falling in a well.

This time the Mohists are expanding on fatalism: 'That there is fate is not fated; to deny that there is fate is to deny fate'. It is harder to reconstruct a problem that is plausibly solved by this analysis.

The algebraic form is abandoned in illustrating the next two breakdowns in parallelism. The first is 'part comprehensive; part not':

- 'Loving people' depends on comprehensively loving people. 'Not loving people' does not.
- 'Rides horses' does not depend on comprehensively riding horses. To have ridden on horses is enough to count as riding horses.

These examples highlight the mass-substantive structure of reasoning about reference. In one phrase the term-reference is implicitly comprehensive, in the other it is not.

Finally, we come to the examples of one 'is this' and one 'not this':

- Fruit of a peach is a peach, fruit of a bramble is not a bramble.
- Asking about a person's illness is asking about the person. Disliking the person's illness is not disliking the person.
- A person's ghost is not the person. Your brother's ghost is your brother. Offering to a ghost is not offering to a person. Offering to your brother's ghost is offering to your brother.
- If the indicated horse's eyes are blind, then we call the horse blind. The horses eyes are large yet we do not call the horse large.
- If the indicated oxen's hairs are brown, then we call the oxen brown. The indicated oxen's hairs are many yet we do not call the oxen many.
- One horse is horse; two horses are horse. Saying 'horses are four footed things' is a case of one horse and four feet, not a pair of horses and four feet. Saying 'horses are partly white' is two horses and some white, not one horse and partly white.

There is no further analysis or summing up. The moral, we assume, is a negative one. Had we treated the first algebraic model as a kind of logic, the Mohists' argument by example would still show that it was invalid; it is not a reliable form in the sense that a true premise formally guarantees a true conclusion. However, the essay is not about sentences and truth at all. It is about whether there are reliable parallels developing from terms to longer (guiding) phrases.

The implicit answer is 'no'. The Mohists offer no constant or consistent principles guiding the construction of longer 'phrases' out of terms even when the terms are consistently applied to external realities. They offer no way to systematically rationalize the conventional patterns of use. They retreat implicitly from Mozi's goal of replacing convention with a constant *dao* (guide). If there is such a 'guide', then it is not a product of any simple projection from term reference. Moral guidance cannot derive from knowledge of natural kinds; it requires conventional,

creative human social activity. Dialectically, the negative result gives ammunition to the Daoists who argue that no constant 'guide' exists (see [Daoist philosophy §2](#)).

7 Gongsun Long: one name, one thing

Against the Mohist background, we can now make some sense of the previously obscure writings of Gongsun Long. The text that bears his name consists of a few dialogues. Each is introduced by a counterintuitive paradox followed by a discussion or chain of reasoning. Graham argued that at least two of these were forged out of misunderstood fragments of the Mohist writings. They apparently copied the phrases after the Canons had been zipped together and the indexing characters mixed into the text.

Two of the remaining dialogues: 'White Horse' and 'Referring and Things', serve as useful examples. They pose difficult puzzles, to which scholars have offered speculative, controversial and mutually inconsistent interpretations. The various interpretations flow partly from different principles used in selecting interpretive theories or 'translation manuals'. A couple of interpretations are given here, serving both to illustrate that point and to allow us to locate and discuss alternative views of Gongsun Long's theory.

In a (perhaps spurious) preface to the dialogues, Gongsun Long explains his motivation. He cites an example of Confucius rectifying names (see §1) and alleges that he is defending that view. Confucians mostly reject the affiliation, but it makes formal sense. If rectifying is to remove ambiguity from guidance, then it requires that only one name in a guidance situation can refer to the object. I either regard the male before me as 'father' or as 'ruler' or as 'person'. Supposing he is all three, if I am to extract guidance from codified rules, I must decide which rule to use. That means deciding which term is relevant to this action situation. The Mohist attack on that policy rejected any one-name-one-thing principle.

The Mohist account of compounding had negative implications from this point of view. Separable or sum compounds, such as 'ox-horse', conform to the one-name-one-thing principle. The scope of each term remains constant when we combine the terms. The combination names a sum of the two. Hard-white compounding, by contrast, violates the principle of strict clarity and consistency in naming. The scopes change when we compound the terms.

Other sources record Gongsun Long as defending two theses: 'separating the inseparable' and 'separating hard-white'. Since Graham argued that the dialogue on 'Hard-White' was forged, we cannot rely on it for an explanation. Still, we can confidently interpret the slogan since 'hard-white' is the Mohist example of an 'inseparable' or 'interpenetrating' compound. To 'separate' them would be to treat them as 'excluding each other' and the compound as a sum compound. Gongsun Long thus objects to the hard-white model.

The phrase 'white horse' selects one term from each type of compound. The White Horse dialogue begins with a question in the canonical analytic form: 'Is "white horse not horse" assertible?' followed by the answer, 'assertible'. The rest is generally agreed to be a discussion between the sophist defending the answer and an 'other' who raises objections. The first defence is that 'white' names a colour and 'horse' names a shape. Shape and colour are different, so a combination of shape and colour is not merely a shape.

The other most cited argument is 'if you ask for a horse, both a black or yellow horse can arrive. If you ask for a white horse, a black horse or yellow horse will not arrive.' This illustrates an argument thread that ' X is not Y ' follows from ' X is different or distinguishable from Y '. The linking theme is that 'white horse' is a combination of two things and this requires that 'white horse not horse' be 'assertible'. A Mohist might respond that 'asking for a white horse' is indeed different from 'asking for a horse', but a white horse is still a horse.

One line of interpretation treats the term 'horse' as referring to the abstract object 'horseness', and thus 'white horse' to 'white-horseness'. The opening sentence thus states the *true* proposition that the two abstract concepts are distinct entities. Since the connected terms are logically singular, the 'not this' represents 'non-identity'. This line of interpretation is motivated by the principle of charity and undermined by the principle of humanity. It makes the puzzling sentence true (by Western lights) but does not explain how the sophist would have had access to the concepts involved. It does not apply to the two terms when used in the supporting arguments, all of which refer to concrete horses.

An alternative interpretation, that 'horse' refers to horse-stuff, can exploit the 'distinct-hence-different' line of

argument and still consistently interpret the concrete references in the rest of the dialogue. If we similarly regard white as the mass-substantive - white stuff - rather than the abstract 'whiteness', we can see a connection to the Mohist theory of compound names. Gongsun Long regards 'intersection' or 'interpenetrating' compounds as contrary to the one-name-one-thing principle. If white horse consists of two names, each should consistently name (scattered) things. Used in combination, their 'naming' should remain consistent. Thus, they should name the sum of the two stuffs and, as in the case of 'ox-horse', 'not-horse' would be assertable of it.

Alternatively, we may either deny that 'white horse' consists of two names or that they are the same names as when used separately. 'White horse' must be thought of as having no essential relation to 'horse' but as a *sui generis* term for a new stuff. We might then say 'white horse is horse' is not analytically true. Its truth is an accident of usage which *could have been otherwise*. Thus, 'white horse not horse' is assertible.

Gongsun Long's argument then becomes a dilemma. Either we regard 'white-horse' as a sum-compound term - in which case the 'ox-horse' result follows - or we regard it as a *sui generis* non-compound name - in which case the conventions of its use could tie it to anything at all; it need not necessarily be horse. The assumption must be that a name is the same only if it has the same scope (names the same thing). Since 'horse' in 'white horse' does not have that scope it is not the same name and its use in the compound constitutes an arbitrary new term.

The other dialogue poses, if possible, even more daunting barriers to interpretation. The first sentence seems to be an explicit contradiction: everything under heaven is *zhi* (pointing), and yet 'pointing' is not 'pointing'. The rest of the dialogue is content-thin and teeters repeatedly on the edge of pure syntactic contradiction. The only nouns are the puzzling *zhi* (pointing) along with *wu* (thing-kind) and *tianxia* (the world). Most interpreters take the issue to be the meaning of 'pointing' (a rare, mild consensus), and most treat it as semantic reference or meaning.

There are reasons for worry: as we saw, the Mohists focused on 'picking out' 'stuff' from its surroundings. There is no evidence of the semantic concept of 'meaning'; 'reference' is a more plausible evolution of that theory. Using 'meaning,' however, makes this dialogue mesh better with the abstract interpretation of the 'White Horse' thesis. Otherwise, it is hard to find any evidence of a sense-reference distinction or find any indication that it is individual objects rather than mereological wholes or types that we 'point to' (see [Sense and reference](#); [Mereology](#)).

One speculative interpretation exploiting the 'reference' interpretation is both philosophically interesting and relevant to issues that emerge in theory of language and metaphysics. Graham treats the crucial first phrase as meaning that although one can refer to each thing, one cannot refer to 'everything' because in that reference one does not refer to one's own act of referring. Zhuangzi later makes a similar argument against assertions of absolute monism: to say everything is one is to have the one and the saying, which makes two. Tempting as it is, the interpretation has little theoretical connection to the White-Horse thesis. Graham treats both dialogues as dealing with the principle that whole is different from the part. The principle in question needs both careful formulation and plausible motivation.

8 Hui Shi: relativism and monism

We have even less direct textual evidence of the philosophical views of Hui Shi. All that remains are ten paradoxical sayings recorded in the [Zhuangzi](#) and some stories of his playing as Zhuangzi's 'debating companion'. Still, we can be confident of his position because plausible motivations for the sayings are more intuitive and are plausibly reflected in Zhuangzi's Daoism (see [Zhuangzi](#)).

Hui Shi focuses on such distinctions as large/small, thick/thin, high/low, south/north and today/yesterday. The common feature is the variability in which term from each pair we can assert of some object. Most of his paradoxes make sense as contrary comments about the same thing, made from different points of view, such as these examples from [Zhuangzi](#) 33:

- Heaven is as low as the earth; mountains are level with marshes.
- The sun from one perspective is in the middle from another declining.
- Natural kinds are from one perspective living and from one dying.
- I go to Yue today and arrive yesterday.

The most important one for theory of language purpose strikes at the Achilles heel of Mohist realism, the

construction of similarity classes:

- The ten thousand thing-kinds are ultimately alike and ultimately different. This is called the great similarity-difference.

As the *Zhuangzi* develops this insight, it suggests that we can find a difference between any two things no matter how alike, and a similarity between any two things no matter how different. So, even if there are objective similarities and differences, they do not justify any particular way of distinguishing between thing kinds. For every category and name, we could have had conventions that just as consistently and with equal ‘world-guidedness’ divide things differently.

However, the list of Hui Shi’s sayings begins and ends with seeming absolutes. He seems to take a classic, so-called Daoist metaphysical view of an undifferentiated single totality:

- The ultimately great which has nothing outside it - call it the Great One!
- The ultimately small has nothing inside it - call it the Small One!
- Universally love the ten thousand thing-kinds; the cosmos is one ‘substantive part’.

The concluding statement echoes the Mohists’ ethical doctrine and one of their technical terms. We do not have Hui Shi’s reasoning, but a plausible explanation is that normally taken by interpreters of Zhuangzi. It is the familiar inference of absolutism from relativism. If all distinctions are relative to some perspective, then we can conclude that reality has no distinctions. All distinctions are false: reality is an undifferentiated total one.

The *Zhuangzi* presentation of Hui Shi’s views notes: ‘He had many perspectives and his library would fill five carts, but his doctrine was self-contradictory and his language did not hit the target: the intent to make sense of things’ (*Zhuangzi* 33). Zhuangzi presumably understood the incoherence of denying distinctions and, if we accept Graham’s speculation about pointing and things, also the notion of an ultimate one - an ‘everything’ concept. Whether or not Gongsun Long rejected the inference, Zhuangzi clearly did. Zhuangzi almost paraphrases Hui Shi:

‘The cosmos and I were born together, the ten-thousand things and I are one.’ Now, having already constructed a ‘one’ is it possible to say something about it? Having already called it a ‘one’ can we fail to say something about it? ‘One’ and saying it make two. Two and one make three and going from here, even a skilled calculator can’t keep up with us, let alone an ordinary man.

(*Zhuangzi* 2)

Zhuangzi displays an immense fondness for Hui Shi alongside a dismissive, almost belittling attitude toward the result of his ‘distinction-dispute’. Traditional accounts have reckoned this as a mystic’s haughty disdain for logic. However, when we grasp that Hui Shi’s doctrines have nothing to do with logic and everything to do with theory of language, a different view of the dynamic emerges. It pairs an erudite, enthusiastic and loquacious but somewhat woolly-minded semantic dilettante (Hui Shi) and a language theorist *par excellence* (Zhuangzi). This is not a case of mysticism versus logic, it is a case of clear versus befuddled theory of language. Zhuangzi enjoyed debating with Hui Shi because he was one of the few with enough learning to be worth refuting, even though he was a relatively easy target for a dialectician of Zhuangzi’s calibre.

9 Zhuangzi: Sceptical perspectivalism

Zhuangzi develops perspectivalism in a more consistent direction. He does not reject language (as perhaps Laozi did). Naturalism, taking the point of view of ‘nature’, does not require abandoning language. Human language, from the empty greetings and small talk to the disputes of philosophers, is another natural ‘noise’. We are all ‘pipes of nature’ (*Zhuangzi* 2).

Zhuangzi grants the Mohist point that language has ‘aboutness’: ‘Language is not blowing breath; there is language for language. That which it languages, however, is radically underdetermined’ (*Zhuangzi* 2). He develops this claim with the aid of Hui Shi’s relativism and his own analysis of the indexicality of all distinctions, and starts by asking to what does ‘this’ or ‘that’ refer? Is there anything that cannot be a ‘this’ or a ‘that’? These terms do not have a rigid, naming relation to an external reality; they trace our changing relations to the realities.

Zhuangzi’s perspectival pluralism is not a version of Western subjectivity. He does not assign any special

perspective to the individual consciousness or internal representations. The kinds of perspective Zhuangzi discusses range from the ways people who speak different languages or accept different moral theories constitute a perspective. At the other extreme, each of us takes different perspectives at different times in our lives - or for that matter, different times of the day.

The Mohists' *ke* (assertible) is also relative to perspective, the changing conventions of usage and principles within disputing factions and schools. Any language that actually is spoken is 'assertible'. Zhuangzi says the appearance of right and wrong in language is a function of elaboration and embellishment of a way of speaking. So the disputes between Mohists and Confucians amount to different elaborated ways of assigning 'is this' and 'not this'.

Zhuangzi calls his 'perspective' on the relativity of language *ming* (clarity). It is a perspective from which we can project backward to an 'axis' of linguistic guidance systems. From that 'axis' there would be no limit to what could be treated as 'is this' or 'not this'; a *dao* (guide) is merely the path one takes from that axis. Thing-kinds are made so by our classifications.

Lacking any limit on possible systems of naming and guiding, we lack any limit on know-how. No matter how much we advance and promote a way of dealing with things, there are things at which we will be deficient. To have any developed perspective is to leave something out. This, however, is not a reason to avoid language and a perspective; it is simply the inevitable result of limitless knowledge and limited lives.

'Sages' project their perspectives and prejudices on 'nature' and 'those who have arrived' know to deem everything as one. Zhuangzi does not recommend we use that attitude. Instead of trying to transcend and abandon the usual or conventional ways of speaking, we should treat them as useful. They enable us to communicate and get things done. That is all it is intelligible to ask of them.

Beyond the usefulness of our language, we don't know the way things are in themselves. We may dub our lack of that metaphysical knowledge as *dao*. To tax our minds by trying to treat everything as 'one' differs from that admission of ignorance only in the emotion that accompanies it. In the end, neither scepticism nor monistic mysticism allows us to say anything about ultimate reality. They merely exhibit different attitudes towards saying... nothing.

10 Xunzi: Confucian conventionalism

The final chapter in Chinese language theory comes in the 'Rectifying Names' chapter of the *Xunzi*. Xunzi focuses on language because he wants to reassert that 'ritual' is the only standard of correct behaviour. He rejected Mencian intuitionism and gleaned insights from Zhuangzi and the dialecticians. The apparent moral he drew was that, since reality cannot be a standard of language correctness, the default standard must be convention.

Appeal to the usage of the sage-kings determines correct name use. The correct account of that usage is a historical tradition, by which Xunzi means the judgment of Confucian scholar-gentlemen. Thus Confucianism is vindicated by the weakness Mozi had exposed. Xunzi then goes on to construct an explicitly conventionalist theory of language which carries political implications.

Xunzi introduced an important clarification. He distinguished between two kinds of 'distinctions': *gui-jian* (noble-base) and *tong-yi* (same-different). The former correspond roughly to value distinctions and the latter to empirical or descriptive distinctions. The latter are the basis for interaction with other cultures, and a king is entitled to change these 'miscellaneous' terms. Even a king, however, cannot change conventional evaluative distinctions (ranks, titles, punishments or anything in the *li* (ritual)). For these we rely on the sage-kings' *dao* (guide) via the scholar-gentlemen's interpretation. Xunzi regards moral terms as conventional 'artifice' arising from thought, not from nature.

Political authorities rectify names for the original Confucian purposes (order and obedience). Xunzi treats the positions and paradoxes of the dialecticians solely in political terms. Philosophy of language causes social instability by undermining the public guiding language. Philosophers confuse the conventional relations of names and make 'is this/not this' unclear. This we must stop: we must have but one standard of terminology. The king, not disputing philosophers and warring schools, will govern introduction of new descriptive terms.

The king should keep three things in mind as he creates names:

- (1) the reason for having names: the reason for having names is coordinating social behaviour and achieving social order. Hence, value terms govern how we assign descriptive terms.
- (2) the basis of classifying as similar and different: we classify by taking the distinctions delivered by sense organs and using them according to the dictates of a heart imbued with the correct evaluative distinctions.
- (3) the essentials of regulating names: the basis of regulating names is social order and the preservation of a stable, traditional scheme of language.

Xunzi's account of classifying similar and different takes a markedly empirical (epistemological) turn. Unlike the Mohists, Xunzi did not rely on claims that reality presented objective similarities and differences. Zhuangzi had argued that human standards of 'is this/not this' were no more natural than the opposing ones of other animals. Taking Zhuangzi's hint, Xunzi focused on human sense reactions to reality. Indeed, no neutral, inter-species ways of distinguishing things as similar and different exist. Still, while the senses of one species work differently from those of another, those of any one species makes similar distinctions.

All humans sense and respond to approximately the same range of natural distinctions (see [Sense-data](#)). The eyes of humans distinguish the same range and bands of colours, the mouth the same classifications of taste, the ears the same range and discriminations of pitch and so on. The shared nature of intra-species distinction-making underwrites the possibility of community and language. Thus, we abandon any appeal to cosmic nature and rely on what is pragmatically possible for humans in achieving natural human goals.

Our language conventionally clusters some sensible differences and ignores others. Historical, conventional standards dictate how it does this. These norms are transmitted into the cultured gentleman's heart when he masters the transmitted, sage-king's scheme of values. The heart rules the sense organs (as it did for [Mencius](#)). It determines what range of sensible discriminants counts as categories for moral purposes. Thus the categories mesh with the moral system of the sage - kings and match the clustering they originated.

It is clear that Xunzi absorbed a good deal of his contemporary theory of language. It is less clear if he understood the arguments and motivation. He deals with the problem of compound terms by ignoring it: 'If a single term is sufficient to convey the intent then use that and otherwise use a compound term.' The intent, presumably, is the conventionally understood intent. Xunzi does accept the Mohist view of names with varying scopes, and disowns the one-name-one-thing ideal. The only important kind of clarity or consistency is the constancy of convention.

Xunzi treats a number of related problems about names in sensitive fashion. He saw that spatial separation was a basis of describing two things of the same kind as two 'stuffs'. He then defines 'change' as being when a thing's spatial position does not change (exhibits characteristic continuity) and its type does. We then treat the thing as the same thing that has changed characteristics. This discussion of metamorphosis is the closest approximation of the classical Western problem of change (see [Change](#)).

Whether or not Xunzi understood the theories behind the paradoxes he criticizes, he clearly did not appreciate them. He exhibits no philosophical fascination with solving conceptual puzzles for their own sake or using them to drive theory. He criticizes paradoxical statements on primarily political grounds, namely the deleterious social effects of asserting their conclusions. Each upsets conventional ways of using terms. Xunzi's solution is political rather intellectual: ban them.

Xunzi classifies the paradoxes into three groups, which vaguely suggests the line of thought leading to them. Each, he argues, violates one of the three insights into names. The reason for naming is coordinating behaviour, so paradoxes which 'use names to confuse names' include the Mohist's claim that 'killing thieves is not killing men'. This uses a theory of names to yield a conclusion that sounds unconventional. So, we forbid saying them.

The second set 'use reality to confuse names', and the central examples are like Hui Shi's relativity paradoxes. These ignore the shared human empirical basis for assigning similarity and difference and use the fact that having different perspectives on reality might lead us to saying unconventional things about size and shape. So, the king will forbid saying them.

The final group use names to confuse reality and includes 'white horse not horse'. Xunzi's analysis does not help resolve the interpretive puzzles about the line of reasoning since it addresses only the pragmatic consequences of allowing such theorizing. His solution, once again, is for the king to avoid and prevent such distracting sophistry.

11 The aftermath: death of philosophy

One of Xunzi's students, **Han Feizi**, was a minor royal in one of the Warring States, who became a central figure of the Legalist school. He had learned a smattering of Chinese theory of language, and he exaggerated the threat of interpretative anarchy to justify repressing philosophy and language creativity. He followed Xunzi's argument that the ruler should enforce uniformity in language but rejected using a scholarly tradition as the norm. His theory of regulation and punishment was based on a crude argument about shape and name which takes us back to the unexplained Confucian notion that names by themselves guide action. An official post is a capsule description of functions (duties) the holder should perform. In the light of recent discoveries, this doctrine appears to be an application of the doctrine of a cult of ruler worship (Huang-Lao) which taught that the 'guide' was in nature and names were embedded in natural shapes.

Legalism became the official doctrine of the repressive Qin empire, which brought the classical period of Chinese philosophy to an abrupt halt (see **Legalist philosophy, Chinese**). In the aftermath, the insights of Chinese theory of language slipped into obscurity. Huang-Lao became the dominant theory surviving during China's philosophical dark age until the importation of Buddhist theory. The early medieval Daoist interpreters argued that we can have names only for things we see. Suppression had worked its magic.

See also: **Chinese philosophy**; **Confucian philosophy, Chinese**; **Daoist philosophy**; **Language, philosophy of**; **Mohist philosophy**; **Mozi**; **Xunzi**; **Zhuangzi**

CHAD HANSEN

References and further reading

- Fung Yu-lan** (1952) *History of Chinese Philosophy*, trans. D. Bodde, Princeton, NJ: Princeton University Press. (A classic account that highlights the abstract interpretation of the 'White Horse' dialogue. Good for general purposes.)
- Graham, A.** (1978) *Later Mohist Logic, Ethics and Science*, Hong Kong: The Chinese University Press, and London: School of Oriental and African Studies. (The only source in English for the Later Mohist text. Difficult: understanding Graham virtually requires knowledge of classical Chinese.)
- Graham, A.** (1981) *Chuang-tzu: The Inner Chapters*, London: Allen & Unwin. (A translation of the *Zhuangzi*, including the summary of Hui Shi's paradoxes.)
- Graham, A.** (1989) *Disputers of the Tao: Philosophical Argument in Ancient China*, La Salle, IL: Open Court. (An easier but less detailed treatment in the context of Graham's account of ancient Chinese thought.)
- Hansen, C.** (1983) *Language and Logic in Ancient China*, Ann Arbor, MI: University of Michigan Press. (Difficult, based on Graham's reconstruction. A philosophical argument for a radically different interpretation of the linguistic doctrines.)
- Hansen, C.** (1992) *A Daoist Theory of Chinese Thought*, New York: Oxford University Press. (An easier and more extended treatment in an account of ancient Chinese philosophy that emphasizes language.)
- Thompson, K.O.** (1995) 'When a "White Horse" is not a "Horse"', *Philosophy East and West* 45 (4): 481-99. (A definitive view of Gongsun Long's most important dialogue.)

Logic in Islamic philosophy

Islamic logic was inspired primarily by Aristotle's logical corpus, the Organon (which according to a late Greek taxonomy also included the Rhetoric and Poetics). Islamic authors were also familiar with some elements in Stoic logic and linguistic theory, and their logical sources included not only Aristotle's own works but also the works of the late Greek Aristotelian commentators, the Isagōgē of Porphyry and the logical writings of Galen. However, most of the logical work of the Islamic philosophers remained squarely within the tradition of Aristotelian logic, and most of their writings in this area were in the form of commentaries on Aristotle.

For the Islamic philosophers, logic included not only the study of formal patterns of inference and their validity but also elements of the philosophy of language and even of epistemology and metaphysics. Because of territorial disputes with the Arabic grammarians, Islamic philosophers were very interested in working out the relationship between logic and language, and they devoted much discussion to the question of the subject matter and aims of logic in relation to reasoning and speech. In the area of formal logical analysis, they elaborated upon the theory of terms, propositions and syllogisms as formulated in Aristotle's Categories, De interpretatione and Prior Analytics. In the spirit of Aristotle, they considered the syllogism to be the form to which all rational argumentation could be reduced, and they regarded syllogistic theory as the focal point of logic. Even poetics was considered as a syllogistic art in some fashion by most of the major Islamic Aristotelians.

Since logic was viewed as an organon or instrument by which to acquire knowledge, logic in the Islamic world also incorporated a general theory of argumentation focused upon epistemological aims. This element of Islamic logic centred upon the theory of demonstration found in Aristotle's Posterior Analytics, since demonstration was considered the ultimate goal sought by logic. Other elements of the theory of argumentation, such as dialectics and rhetoric, were viewed as secondary to demonstration, since it was held that these argument forms produced cognitive states inferior in certitude and stability to demonstration. The philosopher's aim was ultimately to demonstrate necessary and certain truth; the use of dialectical and rhetorical arguments was accounted for as preparatory to demonstration, as defensive of its conclusions, or as aimed at communicating its results to a broader audience.

1 The subject matter and aims of logic

As was their custom in discussing all the branches of philosophy, the Islamic philosophers devoted considerable attention to identifying the subject matter studied by logic and the aims at which logical studies are directed.

Al-Farabi, whose logical and linguistic writings comprise the majority of his philosophical output, epitomizes the approach to logic that is characteristic of Islamic Aristotelianism. In his *Ihsa' al-'ulum (Enumeration of the Sciences)*, he defines logic as an instrumental, rule-based science aimed at directing the intellect towards the truth and safeguarding it from error in its acts of reasoning. He defends the need for such a science of reasoning on the grounds that it is possible for the mind to err in at least some of its acts, for example, in those in which the intelligibles sought are not innate, but are rather attained discursively and empirically 'through reflection and contemplation'. Al-Farabi compares logic to tools such as rulers and compasses, which are used to ensure exactness when we measure physical objects subject to the errors of sensation. Like these tools, logical measures can be employed by their users to verify both their own acts of reasoning and the arguments of others. Indeed, logic is especially useful and important to guide the intellect when it is faced with the need to adjudicate between opposed and conflicting opinions and authorities.

Al-Farabi's view of logic as a rule-based science which governs the mind's operations over intelligibles is repeated in many of his introductory logical works, and it formed the foundation for Ibn Sina's later refinements (see **Ibn Sina**). In the opening chapters of his *al-Madkhal (Introduction)*, the first logical book of his encyclopaedic work *al-Shifa' (Healing)*, Ibn Sina describes the purpose of logic as one of enabling the intellect to acquire 'knowledge of the unknown from the known'. Like al-Farabi, he defends the need for logic by arguing that the innate capacities of reasoning are insufficient to ensure the attainment of this purpose, and thus they require the aid of an art. While there may be some cases in which innate intelligence is sufficient to ensure the attainment of true knowledge, such cases are haphazard at best; he compares them to someone who manages to hit a target on occasion without being a true marksman. The most important and influential innovation that Ibn Sina introduces

into the characterization of logic is his identification of its subject matter as ‘second intentions’ or ‘secondary concepts’, in contrast to ‘first intentions’. This distinction is closely linked in Ibn Sina’s philosophy to his important metaphysical claim that essence or quiddity can be distinguished from existence, and that existence in turn can be considered in either of its two modes: existence in concrete, singular things in the external world; or conceptual existence in one of the soul’s sensible or intellectual faculties (see [Existence](#)).

In *al-Madkhal*, Ibn Sina argues that logic differs from the other sciences because it considers not conceptual existence as such (this would be psychology), but rather the accidents or properties that belong to any quiddity by virtue of its being conceptualized by the mind. These properties, according to Ibn Sina, include such things as essential and accidental predication, being a subject or being a predicate, and being a premise or a syllogism (see [Logical form §1](#)). It is these properties that allow the mind to connect concepts together in order to acquire knowledge of the unknown; they provide the foundation for the rules of reasoning and inference that logic studies. They are moreover formal properties in the sense that, as properties belonging to all concepts in virtue of their mental mode of existence, they are entirely independent of the content of the thought itself; they are indifferent to the intrinsic natures of the quiddities which they serve to link together.

In the *Ilahiyyat (Metaphysics)* of *al-Shifa’*, Ibn Sina introduces the terminology of first and second ‘intentions’ or concepts in order to express the relation between the concepts of these quiddities themselves - which are studied in the theoretical sciences - and the concepts of the states and accidents of their mental existence which logic studies: ‘As you know, the subject matter of logical science is second, intelligible intentions (*al-ma’ani al-ma’qula al-thaniyya*) which are dependent upon the primary intelligible intentions with respect to some property by which they lead from the known to the unknown’ (*Ilahiyyat* Book 1, ch. 2, in Anawati and Zayed 1960: 10-11). For example, the second intentions of ‘being a subject’ and ‘being a predicate’ are studied in logic independently of whatever first intentions function as the subject and predicate terms in a given proposition, for example, ‘human being’ and ‘rational animal’ in the proposition ‘a human being is a rational animal’. The logical second intentions depend upon the first intentions because the first intentions are the conceptual building blocks of the new knowledge which second intentions link together: but logic studies the second intentions in abstraction from whatever particular first intentions the logical relations depend upon in any given case.

2 Logic, language and grammar

The attention that Ibn Sina and al-Farabi devote to the proper characterization of the subject matter of logic stems in part from a concern to distinguish logic from grammar. In the ancient and medieval traditions, the study of logic was closely tied to the philosophical consideration of language (see [Language, medieval theories of; Logic, ancient; Logic, medieval](#)), and for this reason many Arabic grammarians - whose linguistic theories were developed to a high degree of complexity and sophistication - were contemptuous of the philosophers for importing Greek logic, which they saw as a foreign linguistic tradition, into the Arabic milieu. This attitude toward Greek logic is epitomized in a famous debate reported to have taken place in Baghdad in 932 between the grammarian Abu Sa’id al-Sirafi and Abu Bishr Matta, a Syriac Christian who translated some of Aristotle’s works into Arabic and is purported to have been one of al-Farabi’s teachers. The extant account of the debate is heavily biased towards al-Sirafi, who attacks logical formalism and denies the ability of logic to act as a measure of reasoning over and above the innate capacities of the intellect itself. His principal claims are that philosophical logic is nothing but Greek grammar warmed over, that it is inextricably tied to the idiom of the Greek language and that it has nothing to offer speakers of another language such as Arabic.

It is against the background of such attacks that the discussions of the relations between logic, language and grammar by al-Farabi, Ibn Sina and al-Farabi’s pupil Yahya ibn ‘Adi (also a Syriac Christian and translator), are to be understood. Al-Farabi and Yahya both present essentially the same perspective on the relations between logic and language, a moderate perspective which Ibn Sina later rejects. In the *Ihsa’ al-‘ulum*, al-Farabi argues that logic and grammar both have some legitimate interest in language, but whereas grammatical rules primarily govern the use of language, logical rules primarily govern the use of intelligibles:

And this art [of logic] is analogous to the art of grammar, in that the relation of the art of logic to the intellect and the intelligibles is like the relation of the art of grammar to language and expressions. That is, to every rule for expressions which the science of grammar provides us, there is a corresponding [rule] for intelligibles which the science of logic provides us.

(Ihsa' al-'ulum, in Amin 1968: 68)

More precisely, al-Farabi explains that although grammar and logic share a mutual concern with expressions, grammar provides rules that govern the correct use of expressions in a given language, but logic provides rules that govern the use of any language whatsoever in so far as it signifies intelligibles. Thus, logic will have some of the characteristics of a universal grammar, attending to the common features of all languages that reflect their underlying intelligible content. Some linguistic features will be studied in both logic and grammar, but logic will study them as they are common, and grammar in so far as they are idiomatic. On the basis of this comparison with grammar, then, al-Farabi is able to complete his characterization of the subject matter of logic as follows: 'The subject-matters of logic are the things for which [logic] provides the rules, namely, intelligibles in so far as they are signified by expressions, and expressions in so far as they signify intelligibles' (Ihsa' al-'ulum, in Amin 1968: 74).

Like al-Farabi, Yahya ibn 'Adi, in a treatise entitled *Maqala fi tabyin al-fasl bayna sina'atay al-mantiq al-falsafi wa-al-nahw al-'arab* (On the Difference Between Philosophical Logic and Arabic Grammar), makes his case for the independence of logic from grammar based upon the differences between the grammar of a particular nation and the universal science of logic. He argues that the subject matter of grammar is mere expressions (*al-alfaz*), which it studies from the limited perspective of their correct articulation and vocalization according to Arabic conventions. The grammarian is especially concerned with language as an oral phenomenon; the logician alone is properly concerned with 'expressions in so far as they signify meanings' (*al-alfaz al-dalla 'ala al-ma'ani*) (*Maqala fi tabyin*, in Endress 1978: 188). To support this claim, Yahya points out that changing grammatical inflections do not affect the basic signification of a word: if in one sentence a word occurs in the nominative case, with the appropriate vocalization, its signification remains unchanged when it is used in another sentence in the accusative case and with a different vocal ending.

In Ibn Sina's view, however, such accounts of the logician's interest in language and its differences from that of the grammarian did not go far enough. In keeping with his own understanding of logic as the science which studies second intentions, Ibn Sina criticized such earlier attempts to introduce linguistic concerns into the subject matter of logic. In *al-Madkhal*, Ibn Sina labels as 'stupid' those who say that 'the subject matter of logic is speculation concerning expressions in so far as they signify meanings (*ma'ani*)'. However, Ibn Sina does not deny that the logician is sometimes or even often required to consider linguistic matters; his objection is to the inclusion of language as an essential constituent of the subject matter of logic. The logician is only incidentally concerned with language because of the constraints of human thought and the practical exigencies of learning and communication. Ibn Sina goes so far as to claim that, 'if logic could be learned through pure thought so that meanings alone could be attended to in it, then it would dispense entirely with expressions'; but since this is not in fact possible, 'the art of logic is compelled to have some of its parts come to consider the states of expressions' (*al-Madkhal*, in Anawati et al. 1952: 22-3). For Ibn Sina, then, logic is a purely rational art whose purpose is entirely captured by its goal of leading the mind from the known to the unknown; only accidentally and secondarily can it be considered a linguistic art.

3 Conceptualization and assent

While the close links between logic and linguistic studies emerge in the Islamic philosophers' consideration of the subject matter of logic, the links between logic and epistemology come to the fore in the consideration of the divisions within logic and the order of the books within Aristotle's Organon. All the principal Islamic Aristotelians organize their understanding of the divisions of logic around the epistemological couplet of *tasawwur* (conceptualization), and *tasdiq* (assent), which constitute for them the two states of knowledge that logic aims to produce in the intellect.

Conceptualization is the act of the mind by which it grasps singular (though not necessarily simple) essences or quiddities, such as the concept of 'human being'. Assent, by contrast, is the act of the intellect whereby it makes a determinate judgment to which a truth-value can be assigned; in fact, conceptualization is defined in Islamic philosophy principally by contrast with assent. Thus, any act of knowledge that does not entail the assignment of a truth-value to the proposition that corresponds to it will be an act of conceptualization alone, not assent. More specifically, the Islamic philosophers link assent to the affirmation or denial of the existence of the thing

conceived, or to the judgment that it exists in a certain state, with certain properties. Thus, assent presupposes some prior act of conceptualization, although conceptualization does not presuppose assent.

One of the purposes of including a consideration of the *tasawwur-tasdiq* dichotomy in introductory discussions of the purpose of logic is to provide an epistemological foundation for the two focal points of Aristotelian logic, the definition and the syllogism (see [Logical form §1](#)). The purpose of the definition is identified as the production of an act of conceptualization, and the purpose of the syllogism is identified as causing assent to the truth of a proposition. However, since the definition and the syllogism are both considered in the *Prior* and *Posterior Analytics* and the works that come after them in the Organon, the study of the ways of producing conceptualization and assent presupposes as its foundation the study of single terms and propositions in the *Categories* and *De interpretatione*.

4 Predicables, categories and propositions

In keeping with the ancient Greek tradition, the Islamic philosophers considered the books of the Organon to be an ordered series which begins with the study of the signification of simple terms in the *Categories* and then proceeds to the study of propositions in the *De interpretatione*. In addition to these two Aristotelian texts, a work of the Neoplatonist [Porphyry](#), known as the *Isagōgē* (Introduction), was appended to the beginning of this series as an introduction to the study of the *Categories* (see [Aristotle §7](#)). It was concerned with the five predicables: genus, species, difference, property and accident. While all of the Islamic Aristotelians wrote commentaries on the *Isagōgē* and utilized its grouping of the predicables, not all were convinced of its utility as an introduction to Aristotle. [Ibn Rushd](#) openly expresses such doubts in the introduction to his *Talkhis kitab al-maqulat* (*Middle Commentary on the Categories*), where he indicates that his original intention was to omit the *Isagōgē* entirely from his series of middle commentaries on the Organon. At the end of his work on the *Isagōgē* itself, he explains bluntly that he does not believe that Porphyry's text is a helpful introduction to the study of logic and questions whether it is really a logical text at all. His sole reason for completing the commentary, he tells us, was to comply with a request made by his friends.

The logical character of the *Categories* presented a related problem for other Islamic philosophers. In the introduction to his *Sharh al-'ibarah* (*Great Commentary on the De interpretatione*), al-Farabi rehearses some of the controversies inherited from the Greek tradition over the relations between the *Categories* and the *De interpretatione*. As al-Farabi points out, the *De interpretatione* can be understood quite well without a prior knowledge of the *Categories*, and the former work makes no explicit references to the latter. Moreover, the *De interpretatione* is principally concerned with the formal relations amongst propositions, such as contradiction and contrariety, whereas the *Categories* is concerned with the signification or meaning of terms as such. Furthermore, in its opening chapters, the *De interpretatione* considers in formal terms the simple parts of which propositions are composed, that is, the noun and the verb. Despite these concerns, however, al-Farabi opts for the traditional ordering of these books on the grounds that the *Categories* is relevant to the whole of logic, since it studies 'the simplest of the subject matters in which logic actualizes itself'. In his *Falsafa Aristutalis* (*Philosophy of Aristotle*), al-Farabi opts for a similar solution to the logical status of the *Categories*, explaining that it comprises an investigation and classification of 'the instances of being from which the first premises are compounded', which are 'the primary significations of the expressions generally accepted by all' (*Falsafa Aristutalis*, in Mahdi 1969: 82-3).

Al-Farabi's misgivings in both of these texts stem from the largely ontological focus of the Aristotelian *Categories*, which calls into question its placement within the Organon. This concern was echoed later by Ibn Sina, who points out that many of the discussions in the *Categories* would be better placed in metaphysics or psychology, since they pertain to the study of expressions as directly signifying external or mental beings, in other words, to first rather than to second intentions. But since the *Categories* is useful in instructing us how to formulate definitions - which is one of the principal goals of logic - its placement in the Organon can be justified on practical grounds.

Islamic philosophers viewed *De interpretatione* as a study of the composition and truth-values of categorical propositions. Thus al-Farabi, in his great commentary on this text, explains that the term 'interpretation' used in the title of the work means 'complete statement' (*al-qawl al-tamm*). A complete statement, according to al-Farabi, must be one which causes a complete understanding in the mind; in other words, one in which assent occurs along

with conceptualization. This is achieved principally by a simple, predicative, categorical statement (*al-qawl al-jazim al-hamli al-basit*) which affirms or denies a predicate of its subject.

5 Theory of argumentation

For the entire Islamic tradition, the crowning glory of Aristotelian logic is the syllogistic theory outlined in the *Prior* and *Posterior Analytics*, especially the latter. The purpose of logic is to provide the means whereby knowledge is to be acquired, and the most valuable type of knowledge is that which is certain and necessary, that is, knowledge gained according to the paradigm of demonstrative science laid out in the *Posterior Analytics*. This part of logic, in the words of al-Farabi's *Ihsa' al-'ulum*, is 'the strongest and pre-eminent in dignity and authority. Logic seeks its primary aim in this part alone, and the rest of its parts are only for its sake' (*Ihsa' al-'ulum*, in Amin 1968: 89). Even the formal study of the syllogism itself is primarily undertaken for the sake of its employment in demonstrations.

In their formal syllogistic theory, the Islamic Aristotelians mainly follow Aristotle's *Prior Analytics*. While they are aware of the fourth figure traditionally ascribed to Galen, the tendency is to dismiss this figure as superfluous and intuitively implausible, as Ibn Sina does in the seventh method of his *al-Isharat wa-'l-tanbihat (Remarks and Admonitions)*; or to ignore it entirely, as al-Farabi does in his *Kitab al-qiyas (Book on the Syllogism)*. Similarly, the Arabic philosophers knew of the alternative propositional logic of the Stoics and incorporated elements of it in their discussions of conditional or hypothetical (*shartiyyah*) syllogisms (see *Logic, ancient*). However, they did not accept the Stoic inference schemata, nor did they treat conditional connectives as truth-functional, since they did not consider the parts of conditional statements to be complete propositions in their own right. Moreover, for the Islamic logicians 'conditional' was a generic term which included both 'conjunctive' (*al-muttasilat*) conditionals (of the form, 'if... then') and 'disjunctive' (*al-munfasilat*) conditionals (of the form, 'either... or'). Conditional syllogisms of both sorts were viewed as relying upon a process of 'reiteration' or 'repetition' (*istithna*), a term which referred to the repetition of the antecedent or the consequent, or one of the two disjuncts, in so far as it formed the second premise of a syllogism. Thus in the conjunctive conditional syllogism, 'If it is daytime, then it is light; but it is daytime, therefore it is light', 'it is daytime' would be labelled the *mustathna* or reiterated premise, since it is by its restatement that the syllogism reaches its conclusion.

When we turn to the specific application of syllogistic theory to particular types of argumentation, the epistemological concerns of Islamic logic surface once more. In particular, the Islamic philosophers explained the primacy of demonstration, and the ancillary role of dialectical, rhetorical, poetic and sophistical syllogisms, by reference to the epistemic status of the premises used in each type of syllogism, and the type of assent they could produce to the conclusion of the syllogisms in which they were employed. The classification of syllogisms and their premises according to the nature of their assent is found in the logical writings of all the major Islamic philosophers, but the most complete and systematic classification of premises occurs in three of Ibn Sina's works, *al-Burhan (Demonstration)*, in *al-Shifa'*, *al-Najah (Deliverance)* and *al-Isharat wa-'l-tanbihat (Remarks and Admonitions)*. Although these three accounts differ somewhat in the number and variety of the premises listed in each, generally they present a single and consistent theory. Demonstrative syllogisms are composed of premises which necessitate assent and include self-evident first principles as well as sensible, empirically evident propositions. Dialectical syllogisms are based upon generally accepted beliefs (*al-mashhurat*), which are equivalent to the *endoxa* of Aristotle's *Topics*; on premises granted for the purposes of dialectical debate; and in general, on all premises assented to because they are universally accepted by all people, or by people deemed authoritative. Rhetorical syllogisms are similar to dialectical ones, except that they are accepted unreflectively and on the basis of a more limited authority, relative, for example, to a particular group or sect; as such, they are only supposed or presumed to be 'generally-accepted beliefs'. Sophistical premises are those accepted because of some misleading resemblance to another type of premise, and poetic premises are those that produce a motion in the faculty of imagination (*al-takhyil*), not an act of intellectual assent.

The inclusion of rhetorical and poetical syllogisms in this enumeration reflects a common assumption among Islamic philosophers that Aristotle's *Rhetoric* and *Poetics* are parts of his logical *Organon*. This assumption was inherited by the Islamic tradition from the Greek commentators, and it was used by them in part to account for the differences between philosophical and popular modes of discourse and argumentation, particularly in the context of discussions of the relations between philosophy and religion. The Islamic philosophers held that whereas

philosophers rely principally upon demonstrative and dialectical syllogisms, religious leaders and theologians generally use rhetorical and poetical syllogisms to persuade the general populace. Religion is thus viewed as an image or reflection of philosophical, demonstrative truth propounded in language and argument-forms that can be easily understood by the mass of humanity.

The place of dialectic within the theory of argumentation is perhaps the most ambivalent in Islamic logic. While dialectic is seen as inferior to demonstration, its importance for philosophy is none the less recognized. A good example of this is found in al-Farabi's enumeration in his *Kitab al-jadal (Book on Dialectic)* of the ways in which dialectic serves philosophers. According to al-Farabi, dialectic hones argumentative skills, introduces the principles of the special demonstrative sciences, alerts the mind to the self-evident principles of demonstration, helps to develop communicative skills and provides the means for refuting sophistry. Of these five uses, only the fourth is external to the proper aims of philosophy and closer to the tasks usually reserved to theology and religion. The other four pertain to the learning or acquisition of truly philosophical skills, even if they lie outside the strictly demonstrative aims that are the ultimate end of philosophy.

In the case of the theory of demonstration itself, Islamic logicians organized their commentaries on the *Posterior Analytics* around the definition and the demonstrative syllogism as the means by which both conceptualization and assent are most perfectly attained. Al-Farabi's *Kitab al-burhan (Book on Demonstration)* offers an excellent summary of the standard approach taken by Islamic philosophers to theory of demonstration and its epistemological aims. Just as he identified the categorical statement as the embodiment of perfect assent on the propositional level, here al-Farabi identifies demonstrative certitude as complete or perfect assent on the level of syllogistic inference. Moreover, certitude is defined by al-Farabi in terms of what we would now label 'second-order' knowledge:

Certitude is for us to believe, concerning the truth to which we have assented, that it is not possible at all for what we believe about this matter to be different from what we believe it to be; and in addition to this for us to believe, concerning our belief, that another belief is not possible - in the sense that whenever some belief about the first belief is formed, it is impossible for it to be otherwise, and so on *ad infinitum*.

(*Kitab al-burhan*, in al-'Ajam and Fakhry 1986-7, 4: 20)

Certitude requires not just knowledge of a conclusion, *p*, but knowing that we know *p*. This sort of certitude al-Farabi calls 'necessary certitude'. However, he also allows for non-necessary certitude, which holds 'only at a particular time', and thus can be applied to propositions about merely contingent beings: 'Necessary certitude and necessary existence are convertible in entailment, for what is verified as necessarily certain is necessarily existent' (*Kitab al-burhan*, in al-'Ajam and Fakhry 1986-7, 4: 22). While al-Farabi recognizes both of these varieties of certitude to be forms of perfect assent, in his view necessary certitude alone fulfils the strict conditions of Aristotelian demonstration, since it alone will pertain to objects which cannot be other than they are.

Al-Farabi's remarks on the utility of dialectic, combined with his extension of the notion of perfect assent beyond the confines of strict and necessary demonstration, illustrate the overall breadth of the Islamic philosophers' theories of argumentation. Despite their professions of the primacy of the demonstrative paradigm within philosophy, the Islamic Aristotelians recognized a broad range of legitimate and useful argument forms and acknowledged their importance as philosophical tools leading to knowledge of the unknown.

See also: [Aristotelianism in Islamic philosophy](#); [Aristotle](#); [al-Farabi](#); [Ibn Sina](#); [Logical form](#); [Logic, ancient](#); [Logic, medieval](#); [Logic, philosophy of](#); [Meaning in Islamic philosophy](#); [Syntax](#)

DEBORAH L. BLACK

References and further reading

- Abed, S.B.** (1991) *Aristotelian Logic and the Arabic Language in Al-Farabi*, Albany, NY: State University of New York Press. (An excellent consideration of the central issues in al-Farabi's linguistic philosophy.)
- Black, D.L.** (1990) *Logic and Aristotle's 'Rhetoric' and 'Poetics' in Medieval Arabic Philosophy*, Leiden: Brill. (Discusses the interpretation of these Aristotelian texts as works of logic.)
- al-Farabi** (c.870-950) *Ihsa' al-'ulum (Enumeration of the Sciences)*, ed. U. Amin, Cairo: Librairie Anglo-Egyptienne, 3rd edn, 1968. (Al-Farabi's major work, in which he investigates in detail different forms of knowledge.)

- al-Farabi** (c.870-950) *Kitab al-burhan (Book on Demonstration)*, in R. al-‘Ajam and M. Fakhry (eds) *al-Mantiq ‘inda al-Farabi*, Beirut: Dar el-Mashreq, 1986-7, 4 vols. (*Al-Mantiq* is a collection of al-Farabi’s writings on logic.)
- al-Farabi** (c.870-950) *Falsafah Aristutalis (Philosophy of Aristotle)*, trans. M. Mahdi in *Alfarabi’s Philosophy of Plato and Aristotle*, Ithaca, NY: Cornell University Press, 1969. (The *Falsafah Aristutalis* contains al-Farabi’s account of the Organon.)
- al-Farabi** (c.870-950) *Sharh al-‘ibarah (Great Commentary on De interpretatione)*, trans. F.W. Zimmerman, *Al-Farabi’s Commentary and Short Treatise on Aristotle’s ‘De Interpretatione’*, Oxford: Oxford University Press, 1981. (Very learned introduction which sets al-Farabi’s text against the background of the Greek and Arabic traditions in logic.)
- al-Farabi** (c.870-950) Commentary on the *Isagōgē*, ed. and trans. D.M. Dunlop, ‘Al-Farabi’s *Eisagoge*’, *Islamic Quarterly* 3, 1956: 117-38. (Excellent translation of this crucial logical text, given the significance of the *Isagōgē* for this period of logic.)
- al-Farabi** (c.870-950) *Risalah on Logic*, ed. and trans. D.M. Dunlop, ‘Al-Farabi’s Introductory *Risalah on Logic*’, *Islamic Quarterly* 3, 1957: 224-35. (A brief but influential summary of his views on logic.)
- al-Farabi** (c.870-950) Introductory Sections on Logic, ed. and trans. D.M. Dunlop, ‘Al-Farabi’s Introductory Sections on Logic’, *Islamic Quarterly* 2, 1955: 264-82. (Another translation of a section on logic, designed by al-Farabi as a propaedeutic to the study of philosophy itself.)
- al-Farabi** (c.870-950) Paraphrase on the *Categories*, ed. and trans. D.M. Dunlop, ‘Al-Farabi’s Paraphrase of the *Categories* of Aristotle’, *Islamic Quarterly* 4, 1958-9: 168-97; 5, 1958-9: 21-54. (A short summary of what al-Farabi took the main points of the *Categories* to be.)
- al-Farabi** (c.870-950) Commentary on the *Prior Analytics*, trans. N. Rescher, *Al-Farabi’s Short Commentary on Aristotle’s ‘Prior Analytics’*, Pittsburgh, PA: Pittsburgh University Press, 1963. (Another of al-Farabi’s Aristotelian commentaries.)
- Galston, M.** (1981) ‘Al-Farabi on Aristotle’s Theory of Demonstration’, in P. Morewedge (ed.) *Islamic Philosophy and Mysticism*, Delmar, NY: Caravan Books, 23-34. (Clear account of how al-Farabi developed Aristotle’s notion of demonstrative reasoning, the best form of argumentation.)
- Gyekye, D.** (1971) ‘The Terms *Prima intentio* and *Secunda intentio* in Arabic Logic’, *Speculum* 46: 32-48. (Explanation of two key logical terms, linking them to their Greek and Latin equivalents and explaining how they are used in al-Farabi’s logic.)
- Ibn ‘Adi, Yahya** (893-974) *Maqalah fi tabyin al-fasl bayna sina‘atay al-mantiq al-falsafi wa-al-nahw al-‘arab (Treatise on the Difference between the Arts of Philosophical Logic and of Arabic Grammar)*, ed. G. Endress, *Journal of the History of Arabic Science* 2, 1978: 192-81. (Influential work from the early years when the divisions between logic and language were still highly controversial.)
- Ibn Rushd** (c.1170) Short Commentaries on Aristotle, ed. and trans. C.E. Butterworth, *Averroes’ Three Short Commentaries on Aristotle’s ‘Topics’, ‘Rhetoric’, and ‘Poetics’*, Albany, NY: State University of New York Press, 1977. (Ibn Rushd’s short commentaries on areas of thought which he thought were less strongly logical than demonstrative and dialectical reasoning, yet which still embody logical techniques in some form.)
- Ibn Rushd** (c.1174) Middle Commentary on the *Isagōgē*, trans. H.A. Davidson, *Averroes’ Middle Commentary on Porphyry’s ‘Isagoge’ and on Aristotle’s ‘Categoriae’*, Cambridge, MA: Mediaeval Academy of America, 1969. (A middle commentary on a crucial logical text by Porphyry, which was important in linking Ibn Rushd’s use of logical language with that of his Islamic predecessors.)
- Ibn Rushd** (c.1174) Middle Commentaries on *Categories* and *De interpretatione*, trans. C.E. Butterworth, *Averroes’ Middle Commentary on Aristotle’s ‘Categories’ and ‘De interpretatione’*, Princeton, NJ: Princeton University Press, 1983. (Ibn Rushd’s middle commentaries on two Aristotelian texts which are important for what they show of how he developed some of the key terms of his philosophical logic.)
- Ibn Sina** (980-1037) *al-Isharat wa-‘l-tanbihat (Remarks and Admonitions)*, Part One trans. S.C. Inati, *Remarks and Admonitions, Part One: Logic*, Toronto: Pontifical Institute of Mediaeval Studies, 1984. (The logic portion of *al-Isharat wa-‘l-tanbihat*.)
- Ibn Sina** (c.1014-20) *al-Shifa’ (Healing)*, *al-Ilahiyyat (Theology)*, vol. 1 (Books 1-5) ed. G. Anawati and S. Zayed; vol. 2 (Books 6-10) ed. S. Dunya, M.Y. Moussa, and S. Zayed, Cairo: Organisation Générale des Imprimerie Gouvernementales, 1960. (The part of *al-Shifa’* which deals with theological issues.)
- Ibn Sina** (c.1014-20) *al-Shifa’ (Healing)*, *al-Madkhal (Isagōgē)*, ed. G. Anawati, M. El-Khodeiri and F. al-Ahwani, rev. I. Madkour, Cairo: Al-Matba‘ah al-Amiriyyah, 1952. (Ibn Sina’s account of Porphyry’s *Isagōgē*, giving him the opportunity to specify the meaning he gives to logical language.)
- Inati, S.** (1996) ‘Logic’, in S.H. Nasr and O. Leaman, *History of Islamic Philosophy*, London: Routledge, ch. 48,

802-23.(Analysis of the main concepts in Islamic logic, and the role of logic itself in philosophy.)

- Lameer, J.** (1994) *Al-Farabi and Aristotelian Syllogistics: Greek Theory and Islamic Practice*, Leiden: Brill.(A thorough study of al-Farabi's logical writings and their ancient sources.)
- Madkour, I.** (1969) *L'organon d'Aristotle dans le monde arabe (The Organon of Aristotle in the Arab World)*, 2nd edn, Paris: Vrin. (A useful overview of the logic portions of Ibn Sina's *Shifa'*.)
- Mahdi, M.** (1970) 'Language and Logic in Classical Islam', in G.E. von Grunebaum (ed.) *Logic in Classical Islamic Culture*, Wiesbaden: Harrasowitz, 51-83.(An account of the debate between al-Sirafi and Abu Bishr Matta.)
- Margoliouth, D.S.** (1905) 'The Discussion Between Abu Bishr Matta and Abu Sa'id al-Sirafi on the Merits of Logic and Grammar', *Journal of the Royal Asiatic Society*: 79-129(A translation of the famous debate; for a fuller account see also Mahdi 1970.)
- Marmura, M.E.** (1963) *Studies in the History of Arabic Logic*, Pittsburgh, PA: University of Pittsburgh Press.(An uneven collection of studies and translations but with some useful items.)
- Marmura, M.E.** (1975) 'Ghazali's Attitude to the Secular Sciences and Logic', in G.F. Hourani (ed.) *Essays on Islamic Philosophy*, Albany, NY: State University of New York Press.
- Rescher, N.** (1964) *The Development of Arabic Logic*, Pittsburgh, PA: University of Pittsburgh Press.
(Demonstration that in spite of his opposition to philosophy, al-Ghazali was an enthusiastic supporter of logic. Somewhat dated but still useful.)
- Rescher, N.** (1980) 'Avicenna on the Division of the Sciences in the *Isagoge* of his *Shifa'*', *Journal for the History of Arabic Science* 4: 239-50.(Very clear analysis of the text.)
- Sabra, A.I.** (1980) 'Avicenna on the Subject Matter of Logic', *Journal of Philosophy* 77: 757-64.(Excellent analysis of Ibn Sina's approach to logic.)
- Wolfson, H.A.** (1973) 'The Terms *Tasawwur* and *Tasdiq* in Arabic Philosophy and Their Greek, Latin and Hebrew Equivalents', in I. Twersky and G.H. Williams (eds) *Studies in the History and Philosophy of Religion*, Cambridge, MA: Harvard University Press, vol. 1, 478-92.(Important analysis of this critical distinction in Islamic logic, with discussions of its origins in wider philosophy.)
- Zimmermann, F.W.** (1972) 'Some Observations on al-Farabi and the Logical Tradition', in S.M. Stern *et al.* (eds) *Islamic Philosophy and the Classical Tradition*, Oxford: Oxford University Press, 517-46.(Discussion of the role which al-Farabi played in establishing the notion of logic as a separate theoretical inquiry.)

Logic in Japan

'Logic' became an explicit topic in Japanese philosophy only in the twentieth century. Most effort has been directed to developing a dialectical logic in a Hegelian mode rather than a symbolic system. The Japanese term coined for logic in this sense is *ronri* (the 'principles of discourse or argument'). The term *ronrigaku* (*ronri* + -ology) is the more common term for formal, symbolic logic. In the twentieth century two Japanese philosophers, Nishida Kitarō (1870-1945) and Tanabe Hajime (1885-1962), developed distinctive dialectical logics that drew on key assumptions from traditional Japanese thought, but followed a Western style of analysis and articulation.

Nishida's logic of *basho* (place, topos, field) located the contextual premises of empirical and idealist judgments within a trans-judgmental domain called the 'acting-intuiting'. He related this to his logic of the predicate, which rejected the Aristotelian priority of the grammatical subject as signifier of substance, instead making the subject the qualifier of the predicate, the signifier of the event. Tanabe criticized Nishida's system as ahistorical and transcultural. His logic of species gave priority to the specificity of cultural and historical embeddedness, the middle ground between the universal ought and individual freedom.

1 Dialectical logic

To appreciate the development of logic in modern Japan, it is useful to bear in mind three tendencies from traditional Japanese philosophy. First, Japanese thought has tended to emphasize internal rather than external relations. That is, in a relationship the two relatents, *a* and *b*, are generally understood to be not separate entities that have been connected by a relating principle, but to be two overlapping entities: part of *a* is *b* and part of *b* is *a*. Traditional Buddhist philosophy assumes no entity ever exists independently as a discrete substance; it is always in flux and depends for its existence on some other entity (see [Buddhist philosophy, Japanese](#)).

The second point in traditional Japanese thought is the application of internal relations to the part/whole model (see [Mereology](#)). In esoteric, Tendai and Kego Buddhism, the whole has typically not been understood as simply the sum of the parts. Because the whole and parts are internally related, the part (as in a recursive set) always reflects in some way the whole.

The third point about traditional Japanese thought is that even opposites are often understood as being internally related. If, therefore, *a* and not-*a* overlap in some way, that suggests that *a* can only be fully *a* insofar as it contains something of not-*a*. At the least, in defining an entity, it is typically distinguished from and therefore related to its opposite. In East Asian texts, such as the translation of the Indian Buddhist *Perfection of Wisdom* texts, the Sino-Japanese term *soku* is commonly used to express this identity within difference: *a soku not-a*. The term *soku* can be used as a special kind of copula or can be used as a phrase meaning something like 'that is to say'. English translations (not completely successfully) tend to capture this sense by using the Latinate phrases '*a qua not-a*' or '*a sive not-a*' (see [Kyoto School §3](#)).

When Western philosophy entered Japan in the latter half of the nineteenth century, Japanese philosophers became aware that the dominant Western modes of logical expression could not adequately express the premodern tendencies just explained. Among Western systems, Hegel's dialectical logic (see [Hegel, G.W.F.](#)) seemed to most resemble the *soku* way of relating opposites and many Japanese philosophers studied it intensely in the early twentieth century. In the end, however, most Japanese philosophers found it unacceptable. In fact, it seemed almost the mirror image of the kind of dialectical logic sought by the Japanese. First, Hegelian logic is grounded in being, whereas traditional Asian (especially Buddhist and Daoist) philosophy gave more primacy to nonbeing or nothingness. Second, the Hegelian system of dialectic expresses how static beings are transformed into events of becoming. In Japanese thought, however, events are the starting point and logical analysis is used to abstract from it discrete beings within the becoming. Third, the Hegelian dialectic is progressive. It shows the development of thought through time into what it is becoming. In Japanese thought, however, the dialectical process is more commonly used regressively to understand how thought comes about and out of what it is formed. Therefore, although modern Japanese philosophers have continued to use the term 'dialectic' (*benshōhō*), except for Marxists and certain other neo-Hegelians, the term usually refers to a distinctively Japanese logic of dialectic.

Two of the best known dialectical Japanese logics were developed by [Nishida Kitarō](#) (1870-1945) and [Tanabe](#)

Hajime (1885-1962). Both figures were trained in mathematics (in fact, in his younger days, Tanabe was Japan's most renowned philosopher of mathematics and science), well versed in Hegelianism and steeped in Buddhism (Nishida was a Zen Buddhist, Tanabe a Pure Land Buddhist). Both belonged to what is often called the Kyoto School of Japanese philosophy.

2 Nishida's logic of *basho* and logic of the predicate

Nishida attacked Western paradigms of logic from two directions, epistemological and metaphysical/syntactical. When taking the former approach he generally spoke of his logic as the 'logic of *basho*'. *Basho* is an ordinary Japanese word meaning 'place' that can serve as an equivalent for the technical terms *topos* or 'field'. Nishida understood the philosophical construction of reality to take place within three experiential domains or *basho*: empiricism, idealism and what he called the 'acting-intuiting'. The empiricist speaks from out of the '*basho* of being'; that is, the empiricist judges the world of physical entities (see **Empiricism**). The idealist, on the other hand, judges the world of psychological entities, showing their role in the construction of our experience of physical entities (see **Idealism**). This is the realm of consciousness, experience, self and the transcendent ego. Sometimes the Kyoto School refers to this as the '*basho* of relative nothingness', because from the standpoint of the empiricist this psychological world is considered so transparent as to be omitted from the discussion of what is. Relative to empirical reality, it is a place of nothingness.

The third *basho* is supposedly less commonly articulated in the West, although Nishida found suggestive references to it in William **James**, **Bergson**, some neo-Kantians and some religiously or aesthetically oriented philosophers. The third *basho* is where knowing and doing are virtually a single act. It occurs when a person responds spontaneously without deliberation to the information being intuited, or when their intake of information is the same act as going out and gathering information. Nishida calls that immediacy the single event of 'acting-intuiting'. This realm is unknown to the categories of both the empiricist's material world and the idealist's inner world. In fact, in its immediacy and oneness, it cannot be analysed in any systematic way at all. Therefore, it can be called the '*basho* of absolute nothingness' (see **Buddhist concept of emptiness**).

In themselves, these three realms may constitute the contents of kinds of knowing, but Nishida argues there is a formal connection among them: there is a hierarchical 'logic of *basho*'. His analysis of the hierarchy is based in his evaluation of how the *basho* are implicated in each other through the logical analysis of the conditions of judgment. Consider an empirical judgment such as 'The cat is black'. Nishida asks in what place, in what *basho*, one is standing when making that judgment. He argues that for such an empirical judgment to be made, one must have an experience of perceiving that black cat. In other words, one assumes there is a perceiver and a perception that makes the judgment possible. Once we recognize this fact, we realize that judgments within the *basho* of being are logically dependent on experience within the *basho* of relative being, the idealist's world of consciousness and perception. In turn, I may ask about the judgment of 'I see that the cat is black'. That judgment of the *basho* of relative nothingness is itself only possible as a reflection and abstraction out of an experience of the acting-intuiting. So, the *basho* of relative nothingness has as its logical precondition experience within the *basho* of absolute nothingness. The acting-intuiting, being immediate and unitary, is not itself capable of making judgments; yet at the same time, it is the totality out of which reflection can generate the judgments of either idealism or empiricism. For Nishida, it is the universal that becomes concrete through the mediation of the other two *basho*. In this way Nishida's dialectic takes us from the simplest of truths, the empirical judgment, to show regressively its logical basis in the immediacy of absolute nothingness.

Nishida also attacked the Western bias of the ontology of being over nonbeing, of substance over event and of the grammatical subject over the grammatical predicate (see **Being; Substance; Events; Syntax**). Instead of the Aristotelian model in which the sentential subject is correlated with the metaphysical substance and the predicate with the attribute, Nishida proposed a 'logic of the predicate'. In his logic, the primary unit of reality is the event, and only through abstraction do we separate out of that event the sentential subject or the ontological substance. The predicate is a universal whose function becomes concrete and specified through the sentential subject. In doing so, Nishida has in effect made the substance into an attribute of the event. By changing the metaphysical priority from substance to event and the logical or syntactic priority from subject to predicate, Nishida's Buddhism-inspired logic is grounded in change or process rather than stasis or essence. When merged with the threefold model of the logic of *basho*, Nishida attempted to develop a comprehensive philosophical system

including a metaphysics, a phenomenology, an epistemology and an analysis of judgment. His system was at once intelligible to Western analysis, yet simultaneously continuous with the premodern Japanese intellectual tradition.

3 Tanabe's logic of species

Tanabe criticized Nishida's logic of *basho*, proposing two qualifications or fundamental modifications. First, he questioned the implicit assumption that a totally unmediated experience (the acting-intuiting) is possible. Second, he believed Nishida's system was too simplistic in postulating a direct relation between the individual and the universal without the mediation of culture or society. He believed that without taking account of ethnicity and nationality, we cannot truly understand the dialectical development of thought and value through history. The arena of human activity is the social realm, not the unmediated realm of the acting-intuiting, nor the abstracted realm of intellectual philosophical positions like idealism and empiricism.

Tanabe referred to the level of the transcendent universal (that is, not conceptually particularized) totality as the level of 'genus'. Below it is the level of 'species', and below that the level of the 'individual'. In social terms, the logic of genus applies to humankind as a whole, the source of absolute values for ethics and the spiritual basis for the state. The logic of the individual, on the other hand, applies to epistemological judgment and the basis of personal creativity. Between the two lies the logic of species that applies to particularized ethnic, cultural and national values. Without that mediated middle realm, the actual lived experience of human action cannot be explained.

Tanabe began to develop his logic of species during the 1930s, but as he observed developments in Japan during the militarist period, he began to build into the system a more explicit mechanism of self-criticism. He did not feel free to publish his revisions until after the Second World War, however, at which point he emphasized the self-critical moment in his system by developing the philosophical concept of *range*. He himself translated this term as 'metanoia', a philosophical and personal sense of repentance and conversion that leads one to question the absoluteness of any philosophical system one has devised. Essentially, the dynamic of the logic of species is to be the dialectical mediation between the totalitarianism of the genus and the liberalism of individual. From the genus, the species derives the sense of absolute transcendent values applicable to all mankind; from the individual, the species derives the pragmatic concerns of working toward the particular benefit of one's fellow individuals. The nation, indeed the collective or personal activity of any human endeavour, can therefore never speak with the authority of the absolute; on the other hand, no such endeavour can ever be truly individual either, because all expression is enmeshed in the social. Once again, the realm of human activity is not that of fixed being, but that of Buddhist mediation, interdependence and a dialectical tension that deconstructs any system as soon as it makes it claim to be either absolute or individualistic.

See also: [Buddhist philosophy, Japanese](#); [Japanese philosophy](#); [Logic, philosophy of](#); [Nishida Kitarō](#); [Tanabe Hajime](#)

THOMAS P. KASULIS

References and further reading

- Carter, R.E.** (1989) *The Nothingness Beyond God: An Introduction to the Philosophy of Nishida Kitarō*, New York: Paragon House. (An introductory text emphasizing religious and ethical issues; Chapter 2 is a discussion of the logic of *basho*.)
- Dilworth, D.** (trans.) (1970) *Fundamental Problems of Philosophy*, Tokyo: Sophia University Press. (Translation of six of Nishida Kitarō's essays related to his logic of *basho*.)
- Dilworth, D.** (1987) *Last Writings: Nothingness and the Religious Worldview*, Honolulu, HI: University of Hawaii Press. (Translation of Nishida Kitarō's last essays with Dilworth's analysis of Nishida's logic in a comparative context.)
- Dilworth, D. and Satō Taira** (trans.) (1969) 'The Logic of the Species as Dialectics', *Monumenta Nipponica* 24 (3): 273-88. (Translation of Chapter 1 of Tanabe Hajime's *Shu no ronri no benshōhō* (Dialectic of the Logic of Species).)
- Nishida Kitarō** (1911-45) *Nishida Kitarōzenshū* (Collected works of Nishida Kitarō), ed. Watsuji Tetsurō et al., Tokyo: Iwanami Shoten, 1978, 19 vols. (Standard edition of original works.)
- Piovesana, G.K.** (1968) *Contemporary Japanese Philosophical Thought*, New York: St. John's University Press. (Good survey of modern Japanese philosophy including a chapter on Nishida Kitarō and a section on Tanabe)

Hajime's logic of species.)

Tanabe Hajime (1910-61) *Tanabe Hajime zenshū* (*Collected Works of Tanabe Hajime*), ed. Nishitani Keiji *et al.*, Tokyo: Chikuma Shobō, 1963, 15 vols. (Standard edition of original works with vols 6 and 7 being the locus of many of Tanabe's central essays on the logic of species.)

Tanabe Hajime (1946) *Zangedōto shite no tetsugaku*, trans. Takeuchi Yoshinori, *Philosophy as Metanoetics*, Berkeley, CA: University of California Press, 1986. (Translation of one of Tanabe Hajime's major work dealing with metanoia.)

Unno Taitetsu and Heisig, J.W. (eds) (1990) *The Religious Philosophy of Tanabe Hajime*, Berkeley, CA: Asian Humanities Press. (Good collection of essays on Tanabe's philosophy, several of which refer to the logic of species and its differences from Nishida's logic of *basho*.)

Wargo, R.J.J. (1972) *The Logic of Basho and the Concept of Nothingness in the Philosophy of Nishida Kitarō*, Ann Arbor, MI: University Microfilms. (A dissertation, this work has never been surpassed as the best Western treatment of Nishida's logic. It includes a translation of the 'General Summary' from Nishida's *Ippansha no jikakuteki taikei* (*System of Self-consciousness of the Universal*).)

Logic in the 17th and 18th centuries

Logic in the seventeenth century was characterized by attempts to reconcile older viewpoints, such as those of Ramus and Melanchthon, and by criticism of the nature and scope of traditional logic. F. Bacon indicated induction, rather than deduction, as the object of logic, thus opening the way for a logic of the empirical sciences. Descartes proposed to replace the complicated precepts of old logic by simple rules of method. However, even the authors of the Port-Royal Logic, who were influenced by Descartes, could not follow him all the way and continued to teach traditional doctrines, albeit with a new attention to the doctrine of ideas. Other logicians, following Locke, tried to modernize logic by concentrating on an analysis of human cognitive faculties, of the idea-word relation and of other than certain knowledge, thus broadening the scope of logic so as to account for probability. Another suggestion for the improvement of logic came from those who thought that logic should assume mathematics as an example either for its axiomatic-deductive method or for the inventive techniques of algebra. The last of these suggestions prompted research in the area of logical calculi. But this kind of research benefited from the doctrines devised by non-mathematically oriented authors who thus provided the logical framework in which algebraic techniques would be tried. This general background accounts not only for the exceptional logic of Leibniz, but also for some logical calculi worked out in the eighteenth century.

1 Seventeenth-century eclecticism

Seventeenth-century logic was conditioned by the debate of the previous two centuries for and against Aristotelianism and scholasticism, both of which had a quasi-autonomous life until well into the eighteenth century, so that differences of opinion persisted.

However, followers of Aristotle's defender Philip Melanchthon and of the anti-Aristotelian Petrus Ramus merged to form a Philippo-Ramist or Semi-Ramist school (see [Logic, Renaissance §5](#)). They restored the parts of Aristotelian logic left out by Ramus (conversion, the square of opposition, categories, demonstration, fallacies), but also attended to Ramist doctrines (for example, definition and division). The Philippo-Ramist B. Keckermann published a *Systema Logicum* (1600) in which he assigned to logic the aim of being systematic, that is, of being a first science directing the human mind; hence the additional name of Systematics given to his followers, such as J.H. Alsted, author of *Clavis Artis Lullianae (A Key to Lull's Art)* (1609) and *Logicae Systema Harmonicum* (1614), and R. Sanderson, who wrote a *Logicae Artis Compendium* (1618).

Eclecticism also characterized followers of the sixteenth-century logician J. Zabarella, according to whom Aristotle had provided an analytic method of discovery ('*regressus*') for mathematical demonstrations. Many seventeenth-century Zabarellists endorsed Ramist and scholastic doctrines, including F. Burgersdijck, whose *Institutionum logicarum libri duo (Two Books of Logical Institutions, 1626)* became the standard handbook in the Netherlands.

Another eclectic, J. Jungius, was one of the outstanding logicians of the century. He was acquainted with Philippo-Ramist dialectic and with the then recent scholasticism of Fonseca and Suarez. He also had Systematic and Zabarellist teachers and admired the *logistica* of Viète. In his *Logica Hamburgensis* (1635) he dealt with a logic of relations, non-syllogistic inferences, contraposition and special syllogisms.

2 Criticism of the syllogism

However, some of the major philosophers of the seventeenth century persisted in criticizing traditional logic, in particular the syllogism. In his *Instauratio magna* (1620) Francis Bacon, while acknowledging the cogency of the syllogism, observed that it cannot guarantee that the ideas occurring in it are abstracted from reality and defined correctly. Since induction presides over the abstraction and definition of ideas, he indicated induction, rather than deduction, as the object of logical research, thus opening the way for a logic of the empirical sciences.

Descartes repeated the sceptical argument that the syllogism, apart from being sterile, begs the question and has no more claim to certainty than non-inferential truths. As is clear from his *Regulae* (1628), for him the problem of inference was not how to infer (since we do this naturally), but how to infer with certainty. Descartes' solution to this problem rests on an analysis of our scientific knowledge which singles out two cognitive operations: intuition and deduction. Intuition is the basic operation because it makes us certain of simple propositions, whereas

deduction is just a chain of successive intuitions. Having found this solution, Descartes saw little use for a complicated logic: indeed, he considered it detrimental to the natural light of reason. As a substitute he listed, in the *Discours de la méthode* (1637), some rules of method meant as mere guidelines for the cognitive operations whose certainty he claimed to have explained.

3 Logic after Descartes

So limited a scope for logic was by no means universally accepted. For example, H. Fabri [Mosnerius] in the first tome of his *Philosophia* (1646) joined late medieval doctrines to mathematics and devised a combinatory calculus of 576 syllogistic moods in all figures. He also introduced truth tables based on three truth-values: truth, falsity and partial falsity. Later in the century independent research was pursued at a higher level in the *Logica demonstrativa* (1697) of G. Saccheri. He emphasized the demonstrative role of the so-called '*consequentia mirabilis*' showing that 'If not *p* then *p*' entails '*p*'.

But even scholars who accepted Descartes' rules of method as the core of logic had to account for traditional doctrines. This is clear from the *Logica vetus et nova (Old and New Logic)* (1654) of Clauberg; and even more from the *Logica restituta* (1662) of the independent Cartesian A. Geulincx, who revived medieval doctrines, related disjunction, conjunction and negation by means of their truth-conditions and devised a 'logical cube' whose faces represented logical axioms and forms of argumentation.

Old and new doctrines were also adopted in *La Logique ou l'art de penser* (1662), known as 'The Port-Royal Logic', by A. Arnauld and P. Nicole. The authors referred to Descartes' views (the 1664 edition mentions the manuscript of Descartes' *Regulae* (1628), and showed appreciation of Pascal's theory of definition, probability and method (they included a special section on method, imitated by many subsequent logics), and of Augustine's theory of language, from which their attentiveness to the idea-word relation was derived. As for traditional instruction, their syllogistic was more accurate than might be expected of a Cartesian logic. For they devised a combinatory calculus of possible moods with rules to choose the valid ones, and correctly warned against confusing the fourth figure and the first figure with transposed premises. None the less, the Port-Royal logicians agreed with Descartes that deduction has a derivative importance as compared with judgment. And given that they defined judgment as a comparison of ideas, they paid great attention to the ideas themselves. The Port-Royal doctrine of ideas revolves around the notions of comprehension (later, 'intension') and extension. The comprehension of an idea consists of the attributes it comprises in itself that cannot be removed from it without destroying it; the extension of an idea consists of the subjects with which the idea agrees. Comprehension and extension stand in inverse relation: the greater the comprehension of an idea, the smaller its extension, and vice versa. This relation regulated all other relations between ideas and operations (for example, abstraction and composition) performable on them.

The influence of Port-Royal doctrines upon French logic was evident right up to the end of the eighteenth century, for example from the *Logique* (1780) and the *Langue des calculs (Language of Calculi, 1798)* of É.B. de Condillac. In England the Port-Royal Logic was acknowledged even by scholars of a declared Aristotelian disposition, such as J. Wallis, who, in his *Institutio logicae* (1687), rejected Ramist syllogisms with singular terms and classified syllogistic figures competently. H. Aldrich, too, though mostly concerned with his valuable presentation of syllogistic, mentioned Port-Royal doctrines when discussing methodology and the history of logic in his *Artis Logicae Compendium* (1691).

4 The emergence of a logic of cognitive faculties

Towards the end of the seventeenth century, logic began to centre increasingly on an analysis of cognitive faculties. Such an analysis, introduced by Descartes, was carried further in the *Essay concerning Human Understanding* (1690) of John Locke and was the stepping-stone to what has been called 'facultative logic'. In this kind of logic the content of ideas is as important as their form, and equally important are the signs designating ideas. This not only explains why Locke equated logic with semiotics, but also gives an insight into the origins of the importance of language for later philosophy and logic.

The logic of cognitive faculties was promoted by the success of Locke's philosophy which, especially in Germany (for example, in the 1752 *Vernunftlehre (Doctrine of Reason)* of G.F. Meier), was often preferred to Descartes'

because of its avoidance of innate ideas and its applicability to other than certain knowledge (which later brought the study of probability within its scope). The influence of Locke's philosophy is evident in England in both the *Logick* (1725) and the *Improvement of the Mind* (1741) of I. Watts, who proposed a mixture of Locke with Port-Royal, as well as in the *Elements of Logick* (1748) of W. Duncan and the *Logicae compendium* (1759) of F. Hutcheson. Locke also influenced the *Système de réflexions* (1712) of the Swiss J.P. de Crousaz.

5 Logic and mathematics

At almost the same time as the birth of the logic of cognitive faculties, another image of the discipline began to emerge. It was born of the conviction that logic compared unfavourably with mathematics.

Some ascribed the superiority of mathematics to its axiomatic-deductive method. Indeed, Descartes' rules of method were interpreted as recommendations to start with a few simple and acknowledged notions (axioms) and thence to proceed to unknown ones (theorems). The supporters of such an interpretation maintained that the axiomatic-deductive method should replace logic and be the model for the right conduct of understanding. A prime example of this view is the *Medicina mentis* (1687) of von Tschirnhaus, who indicated the mathematical method as a therapy in all fields of knowledge and practical life.

Some attributed the superiority of mathematics to the problem-solving and inventive techniques of algebra. The search for equations relating unknown elements with given elements, exemplified in Descartes' *Géométrie* (1637), was considered the true Cartesian logic and was absorbed into the tradition which viewed mathematics as a universal science of invention (*mathesis universalis*).

This assessment of the merits of algebra outlived the imitation of the mathematical method: the latter was either reduced to a mere synthetic order of exposition, or restricted to mathematics because its wider use was held responsible for the degeneration of Cartesianism into Spinozism. Algebraic tools, on the other hand, were fitted for use in the construction of logical calculi. Despite this, the calculi which were actually or tentatively constructed at the time are not comparable to the later algebra of logic, mainly because they were still bound to the quantitative interpretation of algebra.

6 Leibniz and the emergence of logical calculi

Leibniz's logical calculus is connected with his project of an '*ars characteristica combinatoria*', whose accomplishment required (1) an alphabet of thoughts and appropriate symbols for it, (2) a rational grammar common to all languages, (3) an encyclopedia of the sciences and (4) a universal science by which the contents of all sciences would be ordered systematically and combined so as to discover new truths. This last requirement, in its turn, demanded a logical calculus.

Leibniz could not achieve his entire project, but none the less he wrote several incomplete versions of a logical calculus in papers dating from 1679 to the early 1690s (see Leibniz 1966). Common to all versions - which were related to Leibniz's many logical interests (syllogistic, logic of relations, modality and so on), as well as to his monadology - is the idea that a '*calculus ratiocinator*' is a mechanical procedure for drawing conclusions. It is based on: a universal mathematics similar to algebra; symbols standing for concepts or for propositions; and substitution rules governed by the principle that what is concluded with certain variable letters can be concluded equally with other letters satisfying the same conditions.

For all its exceptionality and complexity Leibniz's calculus represents a development of logic that can also be found in others' work which came soon after. This did not depend on Leibniz's direct influence because his calculus remained unpublished until the nineteenth century. The likely explanation is the existence of certain favourable conditions: the reputation of algebra as a universal means of discovery (see §5); and the existence of doctrines of ideas such as that of Port-Royal and the emerging logic of cognitive faculties (see §§3, 4 above). In this respect Cartesian logic and the logic of cognitive faculties were only apparently disconnected from logical calculi: the former provided the latter with a logical field of application. Indeed Locke, who belittled syllogistic, hoped for an algebra of logic. Admittedly, he did not construct a logical calculus and nor did Hobbes, despite his statement in *De Corpore* (1655) that ratiocination is computation. But an early attempt made by Jakob Bernoulli to establish a *Parallelismus ratiocinii logici et algebraici* (1685), while ending in admitted failure, was demonstrably based on a comparison of algebra with the Port-Royal doctrine of ideas.

7 Logical calculi in the eighteenth century

Between the 1740s and the 1760s progress both in the study of ideas and in algebra prompted a number of logical calculi by German-speaking authors.

The theory of ideas had been much developed in Germany due to the influence of A. Rüdiger. In his *De sensu veri et falsi* (1709) he maintained that the syllogism can be a means of discovery provided that, given a premise, we connect one of its terms with a new concept that stands in one of a set of well-defined idea-relations with the other term. For example, by connecting the predicate-term of a universal affirmative premise with any concept subordinated to the subject-term, we establish enthymematically the conclusion of a 'synthetic syllogism'. In general, Rüdiger's point was that if the syllogism can be used to discover truths, then logic does not have to be given a mathematical formulation in order for it to be inventive.

Taking the contrary viewpoint, Christian Wolff supported a mathematical view of logic maintaining that all mathematical demonstrations consist of chains of syllogisms in the first figure. Therefore in his *Vernünfftige Gedanken von den Kräfften des menschlichen Verstandes (Rational Thoughts on the Powers of the Human Understanding)* (1713) and his *Philosophia rationalis* (1728) he considered it futile to advocate the mathematical method as a substitute for logic or to charge traditional logic with being sterile and nugatory.

In the 1730s a number of authors, strengthened by the confidence in the value of logic characterizing the then dominant Wolffism, resumed logical work even along lines unforeseen by Wolff. In particular, J.P. Reusch in his *Systema logicum* (1734) gave a description of judgments and inferences as based on a set of idea-relations (reminiscent of Rüdiger) and on a principle of substitutivity for ideas.

But Reusch did not take the further step of building a calculus of idea-relations. This step was taken by J.A. Segner in his *Specimen logicae universaliter demonstratae (An Example of a Universally Demonstrated Logic, 1740)*. By using the notion of containment in intension and by an accurate employment of negative terms, he classified, assigned symbols to, and subjected to a set of rules the five idea-relations later known as the Gergonne relations. This produced a rare and beautiful example of a calculus applicable to syllogistic and to some non-syllogistic inferences.

An extensional perspective was chosen by G. Ploucquet in his *Fundamenta philosophiae speculativae* (1759). In his syllogistic calculus all concepts, including predicates, are quantified and there is a single general rule stating that concepts must retain in the conclusion of an argument the quantity they have in the premises.

Many topics were treated by J.H. Lambert in his *Neues Organon* (1764), ranging from probabilistic inference to symbolic knowledge to a theory of truth and error. In it he also represented the *A, E, I, O* propositions and syllogisms extensionally by means of straight lines. Such diagrams were independent of but served the same purpose as the circles used by Euler in his *Lettres à une Princesse d'Allemagne sur divers sujets de physique et de philosophie (Letters to a German Princess on Various Subjects of Physics and Philosophy, 1760-2)*. Circles, lines and other figures had also been used by J.C. Sturm (*Universalia Euclidea*, 1661), J.C. Lange (*Nucleus logicae Weisanae*, 1712), Ploucquet and the then unpublished Leibniz. But Lambert's most interesting contribution to logic is a calculus he worked out in letters and in some of his *Logische und philosophische Abhandlungen (Logical and Philosophical Essays, 1782- 7)* which were published after his death. He based it on concepts, their mutual interrelationships, and on the similarities between logical and algebraic operations. This calculus - which allowed for a tentative logic of relations and was capable of expressing composite propositions - was intensional. This feature was criticized by the Ploucquetian G.J. Holland in a letter (1765) to Lambert in which he sketched his own, extensional calculus. (For a comparison of Lambert's and Ploucquet's calculi see Bök (1766).)

Interest in logical calculi faded in the later part of the eighteenth century due, in all likelihood, to the influence of Kant's ideas on logic. In his *Nova dilucidatio (New Exposition, 1755)* Kant had maintained that Leibniz's *combinatoria* was far from being a genuine '*ars inveniendi*' but apparently did not deny that an *ars inveniendi* based on an algebraic calculus of concepts could be achieved. However, in his *Critique of Pure Reason* (1781) and in his *Logik* (1800), he developed a conception according to which an application of algebra to non-mathematical concepts would have no inventive power. Algebra is synthetic a priori, hence inventive and absolutely certain only on condition that its symbols stand for mathematical concepts. The latter are the only concepts whose objects can

be exhibited a priori in intuition, which gives algebraic proofs heuristic advantages and freedom from error. Kant's separation of logic from mathematics was counterbalanced by his rejection of the conception of a logic of the cognitive faculties that he had originally favoured. His mature view was that pure logic, as opposed to transcendental logic, is a formal science - a view not lightly to be dismissed.

See also: [Logic machines and diagrams](#); [Logical and mathematical terms, glossary of](#); [Port-Royal](#); [Universal language](#)

MIRELLA CAPOZZI

References and further reading

- Aldrich, H.** (1691) *Artis Logicae Compendium*, Oxford. (Aldrich's clear presentation of logic made this book popular until the middle of the nineteenth century.)
- Alsted, J.H.** (1609) *Clavis Artis Lullianae (A Key to Lull's Art)*, Strasbourg; repr. Hildesheim: Olms, 1983. (As well as being a follower of Keckermann, Alsted was interested in Lull's art.)
- Alsted, J.H.** (1614) *Logicae Systema Harmonicum*, Herborn. (As with many Systematics, Alsted also showed some sympathy towards the doctrines of Ramus.)
- Arnould, A. and Nicole, P.** (1662) *La Logique ou l'art de penser*; trans. and ed. J.V. Buroker, *Logic or the Art of Thinking*, Cambridge: Cambridge University Press, 1996. (One of the most influential books in the history of logic.)
- Arndt, H.W.** (1971) *Methodo scientifica pertractatum: Mos geometricus und Kalkülbegriff in der philosophischen Theorienbildung des 17. und 18. Jahrhunderts*, Berlin and New York: De Gruyter. (A good guide to appreciate the importance of mathematics for philosophy and logic in the seventeenth and eighteenth centuries.)
- Ashworth, E.J.** (1967) 'Joachim Jungius (1587-1657) and the Logic of Relations', *Archiv für Geschichte der Philosophie*, 49: 72-85. (A brief outline of Jungius' views on relations.)
- Ashworth, E.J.** (1974) *Language and Logic in the Post-Medieval Period*, Dordrecht and Boston, MA: Reidel. (A classic guide to logic 1500-1650.)
- Auroux, S.** (1993) *La logique des idées*, Montreal, Que., and Paris: Bellarmin-Vrin. (A discussion of the Port-Royal Logic. It also contains a good bibliography of logical works in French 1700-1800.)
- Bacon, F.** (1620) *Instauratio magna (Distributio operis)*, repr. in *The Works of Francis Bacon*, ed. J. Spedding, R.L. Ellis and D.D. Heath, London: Longman, 7 vols, 1857-74; repr. Stuttgart: Frommann (Holzboog), 1962. (This work was a source not only for the logic of empirical sciences but also for later logic textbooks' theory of errors.)
- Bernoulli, J.** (1685) *Parallelismus ratiocinii logici et algebraici*, Basilea; trans. T. Boswell, 'The Brothers James and John Bernoulli on the Parallelism Between Logic and Algebra', *History and Philosophy of Logic* 11: 211-24, 1990. (An early attempt at a logical calculus. This work, which was demonstrably written by Jakob Bernoulli, was an academic dissertation in which Jakob Bernoulli acted as president and his younger brother Johann as respondent, hence the usual but incorrect attribution to 'the brothers Bernoulli'.)
- Blanché, R.** (1970) *La logique et son histoire, d'Aristote à Russell*, Paris: Colin. (An introductory book on the history of logic.)
- Bochenski, I.M.** (1956) *History of Formal Logic*, trans. and ed. I. Thomas, Notre Dame, IN: University of Notre Dame Press, 1961. (An introductory book on the history of logic with several excerpts from original texts.)
- Bök, A.F.** (ed.) (1766) *Sammlung der Schriften, welche den logischen Calcul Herrn Prof. Ploucquets betreffen, mit neuen Zusätzen (A Collection of Writings Concerning Professor Ploucquet's Calculus, With New Supplements)*; repr. and ed. A. Menne, Stuttgart and Bad Cannstatt: Frommann (Holzboog), 1970. (Apart from some texts by Ploucquet giving an improved version of his calculus, this collection includes reprints of texts concerning the debate raised by Ploucquet's work and by its comparison with the work of Lambert; also includes essays by Ploucquet and Lambert stressing the differences between their respective work.)
- Buickerood, J.G.** (1985) 'The Natural History of the Understanding: Locke and the Rise of Facultative Logic in the Eighteenth Century', *History and Philosophy of Logic* 6: 157-90. (This work introduced the term 'facultative logic' in the history of logic.)
- Burgersdijck, F.** (1626) *Institutionum logicarum libri duo (Two Books of Logical Institutions)*, Leyden. (This textbook became the standard logical handbook in the Netherlands after 1635.)
- Capozzi, M.** (1987) 'Kant on Logic, Language and Thought', in D. Buzzetti and M. Ferriani (eds) *Speculative Grammar, Universal Grammar and Philosophical Analysis of Language*, Amsterdam and Philadelphia, PA: Benjamin, 97-147. (Gives an insight in the evolution of Kant's conception of logic.)
- Capozzi, M.** (1994) 'Algebra e logica in Jakob Bernoulli', in *Atti del Congresso 'Logica e filosofia della scienza:*

- problemi e prospettive*, Lucca 7-10 gennaio 1993, Pisa: ETS, 55-74.(Considers one of the first attempts to establish the relations between algebra and logic.)
- Clauberg, J.** (1654) *Logica vetus et nova (Old and New Logic)*, repr. in *Opera omnia philosophica*, ed. J.T. Schalbruch, Amsterdam, 2 vols, 1691; repr. Hildesheim: Olms, 1968.(The work of a German Cartesian who tries to reconcile old and new doctrines for the benefit of students.)
- Condillac, É.B. de** (1780) *Logique ou les premiers développements de l'art de penser*, Paris; repr. in *Oeuvres philosophiques*, vol. 2, ed. G. Le Roy, Paris: Presses Universitaires de France, 1947-51; repr. and trans. W.R. Albury, *La Logique (Logic)*, New York: Abaris Books, 1979.(De Condillac was among French logicians influenced by Port-Royal doctrines.)
- Condillac, É.B. de** (1798) *La langue des calculs (The Language of Calculi)*, ed. S. Auroux and A.M. Chouillet, Villeneuve d'Ascq: Presses Universitaires de Lille.(Published posthumously.)
- Crousaz, J.P. de** (1712) *Système de réflexions qui peuvent contribuer à la netteté et à l'étendue de nos connaissances, ou Nouvel essai de logique (System of Reflections which Can Contribute to the Neatness and to the Extension of our Knowledge, or New Essay of Logic)*, Amsterdam, 2 vols.(An example of Locke's influence on continental logicians.)
- Daston, L.** (1988) *Classical Probability in the Enlightenment*, Princeton, NJ: Princeton University Press.(An important contribution to the subject.)
- Descartes, R.** (1628) *Regulae ad directionem ingenii*, 1701; trans. in *The Philosophical Writings of Descartes*, vol. 1, ed. J. Cottingham, R. Stoothoff and D. Murdoch, Cambridge: Cambridge University Press, 1984-5.
- Descartes, R.** (1637a) *Discours de la méthode*, trans. 'Discourse on the Method', in *The Philosophical Writings of Descartes*, vol. 1, ed. J. Cottingham, R. Stoothoff and D. Murdoch, Cambridge: Cambridge University Press, 1984-5.(Descartes' contention is that in the search for truth one need follow only four simple precepts, disregarding the great number of those of traditional logic.)
- Descartes, R.** (1637b) *La Géométrie*; repr. and trans. in *The Geometry of René Descartes*, ed. D.E. Smith and M.L. Latham, LaSalle, IL: Open Court, 1925; repr. New York: Dover, 1954.
- Duncan, W.** (1748) *Elements of Logick*, London; repr. Hildesheim: Olms, 1968.(States clearly that the object of logic is the study of the faculties of human understanding and of cognitive processes.)
- Euler, L.** (1760-2) *Lettres à une Princesse d'Allemagne sur divers sujets de physique et de philosophie (Letters to a German Princess on Various Subjects of Physics and Philosophy)*, St. Petersburg, 1768- 72; repr. in *Leonhardi Euleri opera omnia*, Leipzig: Teubner, 1911.(In a few pages the famous mathematician Euler gives a clear representation, by means of circles, of the basic categorical propositions entering simple syllogisms.)
- Fabri, H. [P. Mosnerius]** (1646) *Philosophiae tomus primus... cuncta excerpta ex praelectionibus R.P. Honorati Fabry (First Tome of Philosophy... together with excerpts from R.P. Honoré Fabri's lectures)*, Lyon.(Fabri's calculus of 576 syllogistic moods was discussed by Leibniz in his *De arte characteristica combinatoria*.)
- Gardner, M.** (1983) *Logic Machines and Diagrams*, 2nd edn with foreword by D. Michie, Brighton: Harvester. (An outline of the development of mechanically induced inference.)
- Gaukroger, S.** (1989) *Cartesian Logic: An Essay on Descartes's Conception of Inference*, Oxford: Clarendon Press.(A clear and well-argued introduction to the topic.)
- Geulincx, A.** (1662) *Logica fundamentis suis, a quibus hactenus collapsa fuerat, ... restituta (Logic Restored to its Foundations... From Which it Had Fallen)*; repr. in *Opera philosophica*, ed. J.P.N. Land, Den Haag, 3 vols, 1891-3; repr. and ed. H.J. de Vleeschauwer, Stuttgart and Bad Cannstatt: Frommann (Holzboog), 5 vols, 1968. (It is noteworthy that, contrary to the example of the Port-Royal Logic, published in the same year, Geulincx denied that method is a part of logic.)
- Hacking, I.** (1975) *Why Does Language Matter to Philosophy?*, Cambridge: Cambridge University Press.(The origins of the importance of language for philosophy are traced back to the doctrines of ideas of Locke and Port-Royal.)
- Hamilton, W.** (1860-9) *Lectures on Metaphysics and Logic*, ed. H.L. Mansel and J. Veitch, Edinburgh and London: Blackwood, vol. 2, appendix.(Hamilton gives an account of the logic of many seventeenth- and eighteenth-century authors, with special attention to early attempts at a quantification of predicates.)
- Hobbes, T.** (1655) *Elementa philosophiae: de Corpore*; trans. in *The English Works of Thomas Hobbes*, vol. 1, ed. Sir W. Molesworth, London: John Bohn, 11 vols, 1839; repr. Aalen: Scientia, 1961.(Maintains that reasoning is the same as computation and in particular is the same as adding and subtracting.)
- Holland, G.J.** (1765) 'Letter to Lambert (9 April 1765)', in J.H. Lambert, *Philosophische Schriften*, vol. 9.1, Hildesheim: Olms, 1968, 17-20.(In this letter Holland is critical of Lambert's intensional calculus and sketches his own, extensional one.)
- Hoorman, C.F.A., Jr** (1976) 'A Further Examination of Saccheri's Use of "Consequentia Mirabilis"', *Notre*

- Dame Journal of Formal Logic* 17: 239-47. (This paper takes into account modern literature on the subject.)
- Howell, W.S.** (1956) *Logic and Rhetoric in England, 1500-1700*, Princeton, NJ: Princeton University Press.
- Howell, W.S.** (1971) *Eighteenth Century British Logic and Rhetoric*, Princeton, NJ: Princeton University Press.
(Both Howell's books are somehow outdated, but they are still the only comprehensive introduction to logic in England 1500-1700.)
- Hutcheson, F.** (1759) *Logicae compendium*, Glasgow. (Logic is here proposed as the study of cognitive faculties, but the author also gives a kind of axiomatic presentation of syllogisms.)
- Jungius, J.** (1635-8) *Logica Hamburgensis (Hamburg Logic)*, Hamburg, 4 vols; 2nd edn repr. Hamburg, 1681; ed. R.W. Meyer, Hamburg: Augustin, 1957. (Leibniz held Jungius in the highest esteem as a logician, especially for what he had left unpublished. But Jungius' published work is anyhow commendable for a lucid elaboration of results taken from a rich tradition on scientific proof.)
- Kant, I.** (1755) *Principiorum primorum cognitionis metaphysicae nova dilucidatio (A New Exposition of the First Principles of Metaphysical Knowledge)*; repr. in *Gesammelte Schriften*, Berlin: De Gruyter, 1902-10, vol. 1, 389; trans. F.E. England, in *Kant's Conception of God: A Critical Exposition of its Metaphysical Development Together with a Translation of the 'Nova Dilucidatio'*, London: Allen & Unwin, 1929.
- Kant, I.** (1800) *Logic*, ed. G.B. Jäsche, repr., together with lecture notes taken by Kant's logic students c.1770-90, in *Lectures on Logic*, trans. J.M. Young, Cambridge: Cambridge University Press, 1992. (The only logical work published under Kant's name in his lifetime.)
- Keckermann, B.** (1600) *Systema Logicum*, Hannover; repr. in *Opera omnia*, Genf, 2 vols, 1614; repr. Hildesheim: Olms, forthcoming. (Discusses what makes a system out of a body of doctrines. After Keckermann the term 'system' was commonly used to designate the scientific construction of logic.)
- Kneale, W. and Kneale, M.** (1962) *The Development of Logic*, Oxford: Clarendon Press, 1984. (The most comprehensive introduction to the history of logic.)
- Lambert, J.H.** (1764) *Neues Organon oder Gedanken über die Erforschung und Bezeichnung des Wahren und dessen Unterscheidung vom Irrthum und Schein (New Organon or Thoughts on the Investigation and Designation of the True and its Distinction from Error and Appearance)*, Leipzig, 2 vols; repr. in *Philosophische Schriften*, vols 1-2, ed. H.W. Arndt, Hildesheim: Olms, 1965. (This famous book is divided into four parts: Dianology (or logic proper), Alethiology, Semiotics and Phenomenology.)
- Lambert, J.H.** (1782-7) *Logische und philosophische Abhandlungen (Logical and Philosophical Essays)*, Berlin, 2 vols; repr. in *Philosophische Schriften*, vols 6-7, Hildesheim: Olms, 1967-9. (These essays were published after Lambert's death.)
- Lange, J.C.** (1712) *Nucleus logicae Weisianae... variis addimentis novis... auctus et illustratus (The Nucleus of Weise's Logic... Enlarged and Illustrated by Various New Additions)*, Giessen. (Lange refers to the logic of Christian Weise (1681), but improves on it by using circles to represent not only the usual categorical propositions and syllogisms but also disjunctive, copulative and comparative syllogisms.)
- Leibniz, G.W.** (1966) *Logical Papers: A Selection*, trans. and ed. G.H.R. Parkinson, Oxford: Clarendon Press. (Collected papers on logic from 1679 to the early 1690s.)
- Lewis, C.I.** (1918) *A Survey of Symbolic Logic*, Berkeley, CA: University of California Press; revised edn, New York: Dover, 1960. (An early attempt to give an outline of the development of logic.)
- Locke, J.** (1689) *An Essay Concerning Human Understanding*, ed. P.H. Nidditch, Oxford: Clarendon Press, 1975. (A philosophical classic which had a great impact on the general concept of logic in Europe in the eighteenth century.)
- Meier, G.F.** (1752) *Vernunftlehre (Doctrine of Reason)*, Halle; repr. Hildesheim: Olms, forthcoming; also *Auszug aus der Vernunftlehre (Excerpts from the Doctrine of Reason)*, Halle; repr. in I. Kant, *Gesammelte Schriften*, vol. 16, Berlin: De Gruyter, 1914. (Meier was the first German professor to use Locke's *Essay* as a textbook. He however advocated a form of logic strongly reminiscent of Wolff. Kant used the shorter version as a handbook for his lectures on logic.)
- Nuchelmans, G.** (1983) *Judgement and Proposition: from Descartes to Kant*, Amsterdam: North Holland. (A very important contribution to the subject.)
- Nuchelmans, G.** (1991) *Dilemmatic Arguments: Towards a History of Their Logic and Rhetoric*, Amsterdam: North Holland. (Discusses Geulincx in particular.)
- Ong, W.** (1958) *Ramus, Method, and the Decay of Dialogue*, Cambridge, MA, and London: Harvard University Press. (A general history of Ramism and of sixteenth century logic.)
- Ploucquet, G.** (1759) *Fundamenta philosophiae speculativae (Foundations of Speculative Philosophy)*, Tübingen. (A first formulation of Ploucquet's calculus, which he improved in essays included in the collection edited by Bök (1766).)

- Reusch, J.P.** (1734) *Systema logicum*, Jena; repr. in C. Wolff, *Gesammelte Werke*, §3, vol. 26, Hildesheim: Olms, 1990. (In the second edition of 1741 Reusch refers advanced readers to the work of Segner.)
- Risse, W.** (1964-70) *Die Logik der Neuzeit*, Stuttgart and Bad Cannstatt: Frommann (Holzboog), 2 vols. (The first comprehensive account of logic 1500-1780. It contains extensive quotations from almost all the authors it cites.)
- Risse, W.** (1965) *Bibliographia Logica: Verzeichnis der Druckschriften zur Logik mit Angabe ihre Fundorte*, Hildesheim: Olms, vol. 1, 1472-1800. (An invaluable bibliographical source.)
- Roncaglia, G.** (1996) *Palaestra rationis. Discussioni su natura della copula e modalità nella filosofia 'scolastica' tedesca del XVII secolo*, Florence: Olschki. (Includes a detailed exposition and evaluation of modal theories discussed in German universities 1600-70.)
- Rossi, P.** (1968) *Francis Bacon: from Magic to Science*, trans. S. Rabinovitch, London: Routledge. (Still to be recommended because it connects Bacon with the logical and methodological trends of his time.)
- Rüdiger, A.** (1709) *De sensu veri et falsi libri IV (Four Books about the Sense of True and False)*, Halle; enlarged 2nd edn, 1722; repr. Hildesheim: Olms, forthcoming. (Apart from synthetic syllogistic, the author maintains that mathematics is based on sensation and provides an extensive treatment of non-mathematical probability.)
- Saccheri, G.** (1697) *Logica demonstrativa, theologicis, philosophicis et mathematicis disciplinis accomodata (Demonstrative Logic Suitable for the Theological, Philosophical And Mathematical Disciplines)*, Torino; repr. and ed. W. Risse, Hildesheim: Olms, 1980. (Includes a noteworthy treatment of syllogistic and a careful study of consequences, which include the 'consequentia mirabilis'.)
- Sanderson, R.** (1618) *Logicae Artis Compendium*, ed. with an intro. by E.J. Ashworth, Bologna: CLUEB, 1985. (An important book in the Oxford curriculum. The editor's introduction to the 1985 edition is of particular value also as a general introduction to the logic of the sixteenth century.)
- Schepers, H.** (1959) *Andreas Rüdigers Methodologie und ihre Voraussetzungen*, Kant-Studien, Ergänzungsheft 78, Cologne: Kölner Universitäts-Verlag. (An exemplary account of Rüdiger's philosophy and logic.)
- Scholz, H.** (1931) *Concise History of Logic*, trans. K.F. Leideckeras, New York: Philosophical Library, 1961. (An early attempt to sketch the development of mathematical logic.)
- Segner, J.A.** (1740) *Specimen logicae universaliter demonstratae (An Example of a Universally Demonstrated Logic)*, ed. with an intro. by M. Capozzi, Bologna: CLUEB, 1990. (Segner received little attention from his contemporaries, but is praised by historians of logic for his ability in applying algebra to logic. The editor's introduction to the 1990 edition gives an overall view of logic in Germany in the first half of the eighteenth century.)
- Sturm, J.C.** (1661) *Universalis Euclidea*, Den Haag. (The author works out twelve new syllogistic moods.)
- Styazhkin, N.I.** (1964) *History of Mathematical Logic from Leibniz to Peano*, trans., Cambridge, MA, and London: MIT Press, 1969. (An outline of the development of mathematical logic.)
- Tschirnhaus, E.W. von** (1687) *Medicina mentis, sive tentamen genuinae logicae, in qua disseritur de methodo detegendi incognitas veritates (Medicine for the Mind, or an Essay with Genuine Logic, Investigating the Method for the Discovery of Unknown Truths)*, Amsterdam: Magnus & Rieuwerts; 2nd, revised edn, *Medicina mentis, sive artis inveniendi praecepta generalia (The Medicine of the Mind, or General Precepts of the Art of Discovery)*, Leipzig: Fritsch, 1695; repr. with an intro. by W. Risse, Hildesheim: Olms, 1964. (The author's idea of logic as a method for the discovery of truth and as a therapy against errors is supported by his conviction that logic should follow the example of mathematics but does not produce anything similar to mathematical logic.)
- Venn, J.** (1881) *Symbolic Logic*, London: Macmillan, 2nd edn, 1894; repr. New York: Chelsea, 1979. (A book that has created an interest in the history of logic of the seventeenth and eighteenth centuries.)
- Wallis, J.** (1687) *Institutio logicae*, Oxford. (Wallis was a famous mathematician and one of the most significant English logicians of his time.)
- Watts, I.** (1725) *Logick, or The Right Use of Reason in the Enquiry after Truth*, London. (A very popular book which was reprinted well into the nineteenth century.)
- Watts, I.** (1741) *The Improvement of the Mind, or A Supplement to the Art of Logick*, London. (The large number of reprints of this book was due to the author's ability to put Locke's theory of knowledge in the form of a logical handbook inspired by the Port-Royal Logic.)
- Wolff, C. von** (1713) *Vernünfftige Gedanken von den Kräfften des menschlichen Verstandes*, Halle; repr. in *Gesammelte Werke*, §1, vol. 1, ed. H.W. Arndt, Hildesheim: Olms, 1978; trans. L. Hawes, W. Clarke and R. Collins, *Logic, or Rational Thoughts on the Powers of the Human Understanding, with Their Use and Application in the Knowledge and Search of Truth*, London, 1770. (A little book which gives a noteworthy analysis of the proof of a mathematical theorem in terms of a series of syllogisms in the first figure.)

- Wolff, C. von** (1728) *Philosophia rationalis sive Logica (Rational Philosophy, or Logic)*, Frankfurt and Leipzig, 2nd edn, 1740; repr. in *Gesammelte Werke*, §2, vols 1/2-3, ed. J. École, Hildesheim: Olms, 1983; partial trans. R.J. Blackwell, *Preliminary Discourse on Philosophy in General*, Indianapolis, IN: Bobbs-Merrill, 1963. (An elaborate exposition of logic that became the standard reference book, even of oppositors, not only in Germany but throughout Europe.)
- Yolton, J.** (1986) 'Schoolmen, Logic and Philosophy', in T.H. Aston (ed.) *The History of the University of Oxford*, ed. L.S. Sutherland and L.G. Mitchell, Oxford: Clarendon Press, 565-91. (Especially concerns Locke and his influence on English logicians such as Watts and Duncan.)

Logic in the 19th century

The nineteenth century was one of the most active periods for logic in Western philosophy. It is regarded foremost as being the first time logic became 'symbolic' and 'mathematical'.

There was tremendous diversity and conflict in logic in this period - indeed there were substantial debates over whether there even is a subject called 'formal logic' and over what the most basic logical forms were. There was an explicit discussion in the early part of the century of whether logic should be extensional or intensional, opinion eventually settling on a purely extensional conception. By the end of the century, many of these debates had been resolved or had withered away, and the most widespread conception and practice of logic coalesced in the early twentieth century into a view we now identify with the works of Frege, Russell and Whitehead.

The nineteenth century brought, for the first time, from Boole and Frege, distinct proposals on how to symbolize logic that were both extensively developed and had widespread influence. Boole is correctly regarded as the father of modern symbolic logic. Frege shared two concerns of many nineteenth-century mathematicians - avoiding incorrect derivations and providing rigorous and clear foundations for the infinitesimal and derivative calculus - and therefore sought to develop a very clear notion of mathematical 'proof'. His notation has not been taken up but his influence, especially on the move towards symbolization, has been considerable.

De Morgan was the first logician extensively and symbolically to discuss the logic of relations. However, a systematic algebraic notation for relations was provided only later by Peirce, and developed also by Schröder.

1 Major movements and themes

Among the main topics in nineteenth-century logic were (1) the purpose or scope of logic, (2) the extent to which logic can usefully be symbolized or diagrammed, (3) the organization or codification of basic and derivative logical principles, such as in an axiomatization, (4) the relationship of logic to mathematics, (5) the relationship of types of logic, such as deductive and inductive, to each other, and to other ideal forms of reasoning, such as the scientific method, (6) how far logic should depart from the account of quantification in Aristotle's theory of the categorical syllogism, (7) the relationship of types of deductive logic, especially of categorical to propositional logic (usually called 'hypothetical logic' in the nineteenth century), and (8) the extent to which logic needs to treat relations in a special way.

There was little discussion of modal logics; of what we now call 'deviant' logic, such as discussions of nonstandard connectives (including the non-material conditional); or of the relationship of logic and natural language. There was an explicit discussion only in the early part of the nineteenth century (and in [Schröder 1890-1905, vol. 1](#)) of whether logic should be 'extensional' or 'intensional', opinion eventually settling on a purely extensional conception. Extensional logic treats terms only according to the concrete entities to which they refer; intensional logic treats terms according to their 'meaning', variously called intension, comprehension or connotation (see [Intensional logics §1](#)). C.I. Lewis (1918) advanced the thesis that the extensive symbolic development of logic required this extensional conception.

An already existing English (nominalistic) penchant for a logic of individual, concrete things was further solidified in works by Whately, Boole, Mill, De Morgan, Venn and Peirce; W.S. Jevons (1864) is unusual for his intensional theory of logic. German logic, owing to its often-noted idealist tendencies, had traditionally gravitated towards the approach that logic deals with concepts, not with 'things'. But works by Hermann Grassmann (1844) and Schröder (1877, and the monumental [Vorlesungen](#) of 1890-1905) had reversed this course. The works of Frege fall squarely within the older German intensional tradition, but this feature was virtually ignored by his English popularizers, such as Russell, and was later conflated with the firmly extensionalist, set-theoretic views of Dedekind and Cantor that arose contemporaneously in mathematics.

2 Purpose and scope of logic

Early modern philosophy, from the Renaissance to the late eighteenth century, had bequeathed numerous candidates for what was to be identified as the purpose or goal of 'logic'. The Renaissance Reform movement in logic, begun by Ramus and continuing through the Port-Royal logicians, modified and simplified the Aristotelian

conception of logic, and augmented it with a study of ‘method’ - which included inductive logic, for example. The English textbook movement, associated with works by figures such as Isaac Watts in the eighteenth, and Richard Whately and J.S. Mill in the nineteenth centuries, adopted standard Reformist topics, but were less interested in a philosophical theory of logic than in improving the reasoning of their student-readers. (Many twentieth-century textbooks unwittingly follow the Reformist topics and the pedagogical approach.)

Several issues in the very characterization of logic itself emerge. Was logic to cover all forms of reasoning, or only deductive reasoning? Was logic to be a theoretical account of the patterns of valid inference, or a practical, ‘how-to’ manual to help readers reason better or identify the faulty reasoning of others? The Reform and English-textbook traditions included treatments of deductive and inductive logic, together with other types of reasoning (analogy, causation, or scientific reasoning). This broad scope, together with an overriding pedagogical purpose, usually kept writing at a fairly superficial level, simply summarizing earlier views. The influential and subtle text of Richard Whately (1826) and the popular text of Mill (1843) fall squarely within this tradition. Even Augustus De Morgan (1847) treated ‘the Calculus of Inference, Necessary and Probable’. The works of George Boole and Charles S. Peirce are also devoted to both deductive and inductive logics. (Only Peirce seems to have had a fully integrated and original theory of reasoning of all sorts; other authors simply followed the Reform syllabus.) Presumably because of a long-standing interest in a priori sentences and forms of inference going back to Leibniz, German logic was always focused exclusively on deductive logic. This included authors as diverse as Kant, Moritz Drobisch, Ernst Schröder and Frege.

Kant’s views had considerable impact on German logic in the early nineteenth century and sharpened the tension between traditional, narrower conceptions of logic and a ‘psychological’ logic about the way humans must reason (see Kant 1800). Kant’s conception of logic as the science of a priori judgments of all sorts, popularized by William Hamilton, led many English writers (Boole, Alexander Bain, Mill) to speak of the ‘laws of thought’. In Germany, psychological theories and philosophical theories of logic were more thoroughly mixed by Wilhelm Wundt, Wilhelm Schuppe, Otto Külpe and - under the additional influence of Hegel - Christoff Sigwart and R.H. Lotze (and F.H. Bradley and B. Bosanquet) and many other psychologist-philosophers. These unholy mixtures brought fulminations from Frege that have been widely repeated in the twentieth century, often to suggest a fatal empirical taint to all of nineteenth-century logic. They are, however, rather ‘laws’ in the sense of how human beings ‘must’ (Kant) or ‘should’ (Peirce) reason; this is at worst a transcendental or normative theory of judgment and reasoning, rather than being simply empirical psychology.

Georg Hegel (1812) attacked what had been the often underdescribed guiding concept of most previous theories of logic, namely the idea of ‘logical form’, arguing that the distinction between ‘form’ and ‘content’ (of judgments or thoughts) was illegitimate. His own conception of logic defies simple summary, but vastly extended the Kantian conception. The extensions and critiques of Kant and Hegel drove early nineteenth-century German logic away from traditional logic. German logic as a philosophical discipline survived in the works of several anti-Kantian (and anti-Hegelian) logicians, among them Solomon Maimon, Moritz Drobisch (1838) and Bernard Bolzano, and was renewed through the later historical reappraisal of Leibniz and Aristotle by F.A. Trendelenburg (1847-67). Its true resurrection was, however, almost entirely due to professional mathematicians - Schröder and Frege, then Hilbert - who had little understanding of, or sympathy with, traditional philosophical logic.

3 Logic and mathematics: Boole

The idea of an ‘algebra of logic’ is a very old one, going back at least to observations by Jakob Bernoulli in 1685 (see Logic in the 17th and 18th centuries), with developed theories in unpublished manuscripts by Leibniz, and in published works by Lambert, Ploucquet, Euler and Grassmann (1844). George Boole then could modestly be credited only with the ‘rediscovery’ of the algebra of logic (1847, 1854). (Boole was apparently unaware of all these precedents, but was perhaps influenced by English views that were creeping towards a notion of an ‘abstract’ algebra, such as George Peacock’s.) Nevertheless, his was an extensional theory of logic, was more developed than previous theories, occurred at a time when logic was ripe for a new and more mathematical approach and so had a sustained impact (direct or indirect) on all later symbolic theories (except Frege’s). Boole is for these reasons correctly counted as the founder of modern symbolic logic.

However, Boole’s logic had four features that were widely perceived as shortcomings. First, it had no account of relations. C.S. Peirce (1867-90) merged Boole’s theory with De Morgan’s work on relations to develop the

‘algebra of relations’ (see §5 below). Second, Boole treated propositional logic as a different interpretation of terms (as classes of ‘occasions’ or of ‘times’, rather than of concrete things). This reduction seemed unproblematic to Booleans but became suspect as Hugh MacColl (in a series of articles, ‘The Calculus of Equivalent Statements’, 1877-80) and Frege drew attention to what we now think of as propositional logic. Third, particular statements (using ‘Some...’, that is, *I* or *O* propositions) were enduringly problematic, and eventually pointed the way towards the generally inadequate handling of quantification by Booleans. Boole would indicate that some *A*’s are *B*’s by writing $A = \vee B$, where ‘ $\vee$ ’ can be thought of as an operator that forms an arbitrary non-empty subset. Peirce and Schröder eventually introduced quantifier-like class operators, Π and Σ , to form classes with, respectively, all individuals and some individuals having a given logical attribute. The intersection of two classes, which we indicate by $A \cap B$, might then be written as $\Pi i A i \times B i$ - the class of all *i*’s that are both *A* and *B* - as well as by the simpler $A \times B$.

Finally, Boole’s symbol, ‘+’, as described unambiguously in his more influential work of 1854, which can mean class union or propositional disjunction, was notoriously non-inclusive: ‘ $A + B$ ’ is interpretable only when *A* and *B* are mutually exclusive. Boole was not without his reasons for this troublesome requirement - it greatly simplifies the calculation of joint probabilities in inductive logic, permits an inverse operation to ‘+’, and gets rid of the odd ‘ $1 + 1 = 1$ ’ (Lambert’s reasoning as well) - but as a feature of deductive logic, it was widely regarded as merely cumbersome. An inclusive interpretation was proposed by Jevons and Peirce in 1867, and this was widely adopted among Booleans. John Venn (1881) held on to the exclusive interpretation - presumably because of his joint concern for inductive logic.

Boole’s earlier work (1847) is in some ways a remarkable, subtle and little-understood work: it does not discuss ‘classes’ and instead gives a theory of what Boole calls ‘elective symbols’. These behave like subset-forming operators on the universe, 1.

Although Boole was paid homage by Peirce and Schröder as the founder (for Schröder, with Grassmann) of the algebra of logic, there are nevertheless respects in which they moved considerably beyond him. They thoroughly revised his handling of quantification, increasingly using quantifier-like operators. They were more interested in the algebra of relations, regarding the simple algebra of logic as somewhat trivial. And finally, they eventually moved away from Boole’s ‘equational’ style of presentation. That is, rather than using a primary symbol of class identity, ‘=’, they used a symbol of class subsumption (‘subset’): Peirce’s ‘ $\prec$ ’ and Schröder’s ‘ $\subseteq$ ’.

A deeper difficulty was with Boole’s apparent reasoning why we should want to turn logic into algebra at all. He seems to have had the idea that doing so would make logic-learning easier and, since many people had algebraic symbol-manipulating abilities, logical reasoning could then rely upon these. He did not articulate deep or philosophical reasons why logic really ‘is’ an algebra; it was convenient and helpful to think this way. (In largely unpublished works of the early 1880s) Frege extensively criticized the underlying motivations for this connection between algebra and logic as superficial and philosophically casual.

The works of Peirce and Schröder in the late nineteenth century moved in a direction compatible with Frege’s critiques. Namely, they became less and less algebraic: no longer equational, and with the introduction of quantifiers, they contained variable-binding operators that disallow context-free algebraic-style substitutions. They also ceased pointing out parallels between logic and numerical algebra, and substantive uses of ‘1’ and ‘0’ disappeared.

4 Logic and mathematics: Frege

Frege speaks only briefly of his conception of the purpose of logic: it is to use a ‘*Begriffsschrift*’ (‘conceptual notation’ - Trendelenburg’s translation of a concept used in discussing Leibniz’s twin goals of a ‘*calculus ratiocinator*’ and a ‘*characteristica universalis*’). That is, Frege’s goal in his notation is to display in a perspicuous way the relationships between concepts and propositions (and presumably not merely those that are usefully similar to relationships in mathematics). The goal of the whole of logic is to demonstrate the correctness of deductions without gaps between premises and conclusions, using acknowledged formal and precise rules of inference. Frege’s focus seems initially to have been on determining the correctness (and hence non-synthetic, a priori nature) of mathematical proofs, rather than examining reasoning of all sorts, as had traditionally been the subject of logic. Logic before Frege had not devoted itself only, or even extensively, to reasoning in mathematics;

mathematicians were thought to be those least in need of help. Frege seems to have shared two concerns of many nineteenth-century mathematicians: avoiding incorrect derivations, like many faulty ‘proofs’ of the parallel postulate from other axioms and postulates in Euclidean geometry; and providing rigorous and clear foundations for the infinitesimal and derivative calculus (‘analysis’ - see [Analysis, philosophical issues in §§1-2](#)). His solution was, in essence, to develop a very clear notion of what counts as a mathematical ‘proof’.

One might then guess that Frege was importantly involved in the axiomatic movements of the late nineteenth century, which first occurred in geometry (Hilbert and Veblen), then extended to the axiomatization of number theory (Peirce, Dedekind and Peano), set theory (Zermelo) and logic itself (R.A. Bernstein and Hilbert). This is not precisely correct, for Frege does not have a theory of axiomatics; certainly he had a general desire for care and precision, and developed reasons for this care and clarity, but his own demonstrations are neither more nor less likely to be flawed than those of Boole, Peirce, or Schröder, and are not axiomatic in any definite sense. Furthermore, Frege’s notational system was so unusual and cumbersome that even Peano insisted that Frege’s formulas be translated into his own notation before he would consider them. Our present notation for quantificational logic derives primarily from that of Whitehead and Russell’s *Principia Mathematica*, which they in turn borrowed from Peano (see [Peano 1973](#)) - and these symbolisms were taken from original works by Peirce and Schröder’s notational devices.

Frege’s contribution lay elsewhere. He developed the first theory of quantification. The Boolean quantifiers of Peirce and Schröder were typically only class abstraction operators, with implicit rules given for their use. Frege also gave a very careful account of propositional logic (using notational devices for the material conditional and for negation) and made it the core of his theory - whereas for Aristotelians and Booleans, propositional logic was dependent on the logic of categorical statements, understood as a theory of classes. Furthermore, in later works, Frege was to use this theory to develop an account of the nature of numbers ([1893, 1903](#)) that was to have an enormous impact on Russell, and on the philosophy of mathematics. Finally, Frege contributed a great many concepts that have become part of the philosophy of logic: what we now call predicates and propositional functions; the use of ‘functions’ in logic; the distinction between sense and reference (see [Sense and reference §§1-2](#)); and many others.

Logic can be ‘mathematical’ in several distinct senses. It can simply use mathematical symbols and rules governing them; this is the method of Boole. It can be developed in the methodical, non-circular, clear way that was epitomized in Euclid’s axioms and postulates for geometry, but had rarely been applied to other branches of mathematics, or to logic until the very late nineteenth century. It can be an independent discipline that is applied to mathematics, to give an analysis of concepts such as number or proof; this is closer to Frege’s methodology. Finally, it can use mathematical techniques - of diagrams, demonstrations and proofs, and careful descriptions of mathematical structures. This is the modern and popular sense in which logic is sometimes said to be ‘mathematical’. Peirce and Schröder might have understood and appreciated a claim, for example, that propositional logic defines a distributive lattice - a certain abstract algebraic structure - but Boole probably would not have understood such a claim and Frege would not have thought it insightful.

5 Relations

De Morgan is famous in logic for an often-repeated sentence he never wrote, roughly to the effect that all of Aristotelian logic was helpless to show the validity of the inference ‘Since every horse is an animal, every head of a horse is the head of an animal’. The spirit is De Morgan’s (except that he was surprisingly sympathetic to Aristotelian logic), but the example was Jevons’. The greatest logical and philosophical minds had long been aware of difficulties with relations, although they had often treated relational arguments as a marginal phenomenon. Plato and Aristotle were keenly aware of issues with regard to relations, and the medievals developed many theories of them (in supposition theory and in the theory of ‘oblique’ syllogisms - see [Logic, medieval §§5, 7](#)). Joachim Jung, the Port-Royal logicians and Leibniz were all aware of this literature - as were De Morgan and Peirce.

With the exception of the remarkable treatment of functions by Lambert in the mid-eighteenth century, De Morgan ([1860](#)) deserves credit for being the first logician extensively and symbolically to discuss the logic of relations. He briefly surveys the history of relations (starting with Aristotle), argues for the importance of relations in common thought, and for the centrality of relations in mathematics. He argues that traditional logic had only treated the

relations ‘identity’ and ‘non-identity’. De Morgan emphasizes the formal properties of identity that form the basis of inferential operations, namely transitivity and ‘convertibility’ (what we would now call symmetry or commutativity). He then proposes a notation for relations (or at least for identity), but demurs from a systematic algebraic notation, later to be provided by Peirce.

Peirce first made his proposals in a rambling essay of 1870. They were presented in a more succinct manner in papers of 1882 and 1883. He published his most successful axiomatic-like presentation of the non-relational and relational algebra of logic (extensively using the quantifier-like Σ and Π operators) in 1885 (see Peirce 1867-90). His logic of relations, unlike many logical proposals, was woven into his more general philosophical system, into his philosophy of mathematics, and into his theory of infinity. (Peirce, following a hint from De Morgan, defines infinite classes as those in which certain relational arguments are invalid, although they are valid for finite ones: this is the theory of the ‘transposed’ syllogism.) Peirce used his system extensively, proved a number of results in the logic of relations, invented a number of technical tools (such as ‘prenex’ form, in which all quantifiers occur first and with global scope), and suggested Church’s theorem: that there is no general mechanical routine for determining the validity or invalidity of an argument once relations and quantifiers are permitted.

In 1895, in the third volume of his *Vorlesungen*, Schröder presented the algebraic theory of relations in a lengthy but highly organized fashion, including the proofs of hundreds of theorems. He uses Peirce’s notation and theory, adding some notational variants - he is notably careful about the distinct universes of discourse (Boole’s ‘1’) for two- and higher-place relations, and, more fastidious still, proposes distinct empty classes (Boole’s ‘0’). Schröder uses Peirce’s quantifier-like Σ and Π expressions, but presents both relational logic (1895: vol. 3) and non-relational logic (1890: vol. 1) in a less axiomatic-like way than Peirce had in 1885. He connects the logic of relations to various matrix, numerical and geometric representations, sometimes using these as what we would now describe as ‘models’. Bertrand Russell in 1910, already familiar with Frege’s and especially Peano’s notations, pronounced the algebra of logic ‘cumbersome’.

The Peirce-Schröder notation and theory with relations was used, mainly in German-language logical work, until about 1920 (and in Skolem’s work for slightly longer). An overview is presented in English in Whitehead’s work of 1897, but otherwise no translation of Schröder’s *Vorlesungen* has ever appeared in English or French. (Although reprinted in 1966, it runs to over 2,000 pages and lacks an index.) Peirce’s work was buried in American mathematics journals of the 1870s and 1880s, at least until the publication of his *Collected Papers* from 1931 on. He published no book summarizing his logic. Consequently, since the 1920s, the algebra and logic of relations as a distinct field has been known only to scholars of the history of logic and mathematics, or to those fluent in German. Interestingly, no conclusive proof has ever been printed that the algebra of logic, particularly when enriched with its quantifier-like expressions or with relations, is inadequate in any technical respects. (Tarski claimed in 1941 that there were formulas of first-order predicate logic it could not express, but did not offer a complete demonstration.)

Frege stands curiously outside of the history of attempts to deal with the logic of relations. This is because he seems to have been completely unaware of the history of the logic of relations, and because his own theory was developed so independently of the other efforts to address the problem (that is, he was unaware of the works of De Morgan and Peirce, while Schröder’s work on relations appeared after Frege’s system was already formulated). His own theory, even in the early *Begriffsschrift* (1879), can obviously express everything that the algebraic theories could. Propositional functions (predicates) can have one or more argument places (1879: §10), but Frege does not draw attention to the difference between one-place functions (monadic predicates such as ‘is a cat’) and those with two or more places relational predicates, such as ‘is smaller than’). He seems unaware of or uninterested in the different logical and philosophical behaviour of properties and relations.

6 Main discoveries of nineteenth-century logic

In terms of successful influences, systematic symbolizations comprise the most obvious logical heritage of the nineteenth century. This is a tradition that clearly began with Boole, was modified and extended by Peirce (and to a lesser extent by Schröder), and then slightly reformed by Peano. Frege had little influence on notation now in common use, but considerable influence on the impetus to use symbolization. The linking of logic with mathematics had enormous impact on twentieth-century logic, but there are many distinct such links: the exact connections seen by De Morgan, Boole, Peirce and Frege are thoroughly different, ranging from the view that both

logic and mathematics can and should use similarly behaving symbolic systems, to views that mathematics ‘is’ logic (Frege, and perhaps Peano), or that logic ‘is’ mathematics (Boole, De Morgan, certainly Peirce). A further influence was the use of variable-binding operators, like those seen in mathematics (especially in the calculus). Finally, logic of both the Boolean and Fregean sort regarded a treatment of relational predicates as necessary, and not a marginal phenomenon - although this view is only implicit in Frege.

Both Peirce and Schröder used something quite similar to truth tables, a discovery usually credited to Wittgenstein. They also occasionally used models to show the independence of logical principles, a technique that was rediscovered and applied to logic in the 1920s; models themselves were suggested by the ‘interpretations’ and ‘universes of discourse’ of Boole and De Morgan. A clear notion of the ‘completeness’ of a logical system was absent, although they understood notions of independence and consistency; Peirce had some idea of decidability.

The value of studying nineteenth-century logic today, for other than its historical value, might include: algebraic - or at any rate ‘mathematical’ - structures inherent in logic as a theory of idealized thought or of the world; a reconsideration of the special structures and features to be found in the logic of relations; a perspective of the relationship of logic and mathematics that is closer to De Morgan’s and Peirce’s than to Frege’s or Russell’s; a return to an intensional conception of logic that once dominated logic and was rediscovered by Church in the 1950s and 1960s in the works of Frege; and conceptions of collective entities, and infinite collections, that retreat from the ornate and perplexing ‘sets’ of the twentieth century to something closer to the ‘classes’ of Boole, Peirce and Schröder - or perhaps to the concepts of Frege.

See also: De Morgan, A.; *Logic machines and diagrams*; *Logical and mathematical terms, glossary of*; Mill, J.S. §§2-3

RANDALL R. DIPERT

References and further reading

- Boole, G.** (1847) *The Mathematical Analysis of Logic, Being an Essay Towards a Calculus of Deductive Reasoning*, London: G. Bell & Sons; repr. in *Studies in Logic and Probability*, ed. R. Rhees, London: Watts, and LaSalle, IL: Open Court, 1952; also repr. in *An Investigation of the Laws of Thought*, New York: Dover, 1973. (Boole’s first and in some ways best attempt at a symbolic calculus; far less influential than the *Laws of Thought* (1854) but philosophically far more interesting and exact.)
- Boole, G.** (1854) *An Investigation of the Laws of Thought, on which are Founded the Mathematical Theories of Logic and Probabilities*, London: Walton & Maberley; repr. in *An Investigation of the Laws of Thought*, New York: Dover, 1973. (Boole’s most famous work, and arguably the most influential work in the history of symbolic logic. Some, including Peirce, have nevertheless considered it to be inferior to his (1847 work.)
- Boyer, C.B. and Merzbach, U.C.** (1968) *A History of Mathematics*, New York and London: Wiley; 2nd edn, 1989, repr. 1991. (A standard work in the history of mathematics, with useful discussion of the intertwined histories of abstract algebra and ‘Boolean’ logic, and of the mathematization of logic in the late nineteenth century.)
- De Morgan, A.** (1847) *Formal Logic: Or, The Calculus of Inference, Necessary and Probable*, London: Open Court, 1926. (A careful work from the same year as Boole’s first major logical work; it had less influence than it deserved because of De Morgan’s resistance to algebraic-style symbolizations.)
- De Morgan, A.** (1860) ‘On the Syllogism IV and on the Logic of Relations’, in *On the Syllogism and Other Logical Writings*, ed. P. Heath, New Haven, CT: Yale University Press, 1966. (The first major work solely on the logic of relations, this had an enormous influence on Peirce, and through him on Schröder and others. De Morgan continued to resist an algebraic-style symbolization but was nevertheless highly rigorous.)
- Dipert, R.R.** (1990-1) ‘The Life and Work of Ernst Schröder’ and ‘Individuals and Extensional Logic in Schröder’s *Vorlesungen über die Algebra der Logik*’, in V. Peckhaus (ed.) *Schröder*, special issue of *Modern Logic* 1 (2/3). (The only extensive works in English on Schröder’s life and work.)
- Drobisch, M.** (1838) *Neue Darstellung der Logik nach ihren einfachsten Verhältnissen*, Leipzig: Voss; 3rd revised edn, 1863. (Perhaps the most important and influential logic book in German before Frege, and probably known by him; experiments with notations and diagrams, and shows some independence from the then-dominant Aristotelian and Kantian/Hegelian approaches.)
- Frege, G.** (1879) *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*, Halle: Nebert; trans. ‘*Begriffsschrift*, a Formula Language, Modelled Upon That of Arithmetic, for Pure Thought’, in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge,

- MA: Harvard University Press, 1967, 1-82.(The first major work of logic in a modern, twentieth-century, non-algebraic calculus; a remarkable and much-praised work but with little direct influence on later developments.)
- Frege, G.** (1893, 1903) *Grundgesetze der Arithmetik: begriffsschriftlich abgeleitet*, Jena: Pohle, 2 vols; part 1 of vol. 1 trans. M. Furth, *Basic Laws of Arithmetic: An Exposition of the System*, Berkeley, CA: University of California Press, 1964.(A difficult work of enormous importance and influence, especially on Russell and for logicism: that mathematics was really logic.)
- Grassmann, H.** (1844) *Die lineale Ausdehnungslehre, ein neuer Zweig der Mathematik*, Leipzig: Otto Wigand, 2nd edn, 1862; repr. in *Gesammelte mathematische und physikalische Werke*, ed. F. Engel, New York: Chelsea, 1969; trans. L.C. Kannenberg, 'The Linear Theory of Extension: A New Branch of Mathematics', in *A New Branch of Mathematics: The 'Ausdehnungslehre' of 1844, and Other Works*, La Salle, IL: Open Court, 1995.(A strikingly idiosyncratic symbolic work. Grassmann proposes a theory of all extension [one-dimensional quantities] including logic; predates Boole but was well-known only in the German-speaking world, with the exception of Peano.)
- Hailperin, T.** (1976) *Boole's Logic and Probability: A Critical Exposition from the Standpoint of Contemporary Algebra, Logic and Probability Theory*, Amsterdam and New York: North Holland; 2nd edn, 1986.(Perhaps the best secondary monograph on Boole's logic; unusually sensitive to the integrated aspects of his deductive and inductive logics; possibly the first edition is superior.)
- Hegel, G.W.F.** (1812-16) *Science of Logic*, trans. A.V. Miller, London: Allen & Unwin, 1969.(A work whose major thesis is that there is no distinction between form and matter in judgments, this work had a huge and negative impact on philosophical formal logic in nineteenth-century Germany.)
- Jevons, W.S.** (1864, 1869) 'Pure Logic' and 'The Substitution of Similars', in *Pure Logic and Other Minor Works*, London: Macmillan, 1890.(Jevons' early works were, with a less influential essay of Peirce's (1867), the major English symbolic works to advocate an 'exclusive' approach to the or/union operator.)
- Jevons, W.S.** (1873) *The Principles of Science: A Treatise on Logic and Scientific Method*, New York: Dover, 1958.(An important English logic text, and one of the few that is intensional.)
- Kant, I.** (1781/1787) *Critique of Pure Reason*, trans. N. Kemp Smith, New York: St Martin's Press, and London: Macmillan, 1965.(A major work with little apparent concern for logic; its extension of a priori to include some synthetic judgments, as well as the 'Table of Categories', nevertheless altered nineteenth-century logic.)
- Kant, I.** (1800) *Logic*, ed. G.B. Jäsche, repr., together with lecture notes taken by Kant's logic students c.1770-90, in *Lectures on Logic*, trans. J.M. Young, Cambridge: Cambridge University Press, 1992.(A 'minor' work of Kant, it nevertheless shows the influence of his philosophy and had considerable impact on later German logicians.)
- Kneale, W. and Kneale, M.** (1962) *The Development of Logic*, Oxford: Clarendon Press, 1984.(The standard but now dated work in the history of logic; has usable discussions of the nineteenth century, adequate discussions of major figures such as Boole and Frege, but thin and unphilosophical discussions of other figures and important issues such as relations.)
- Lewis, C.I.** (1918) *A Survey of Symbolic Logic*, doctoral dissertation, Berkeley, CA: University of California Press; revised edn, New York: Dover, 1960.(Quite a dated work in many ways, but one of the few historical works in the field with a single, well-developed, unifying and enlightening thesis: the rise of extensional logic, and its separation from intensional logics.)
- MacHale, D.** (1985) *George Boole: His Life and Work*, Dublin: Boole Press.(The only extensive discussion of Boole's life, including some of his non-logical work - such as his work in calculus.)
- Merrill, D.D.** (1990) *Augustus De Morgan and the Logic of Relations*, Boston, MA: Kluwer.(The only major work on either De Morgan or the early history of relations.)
- Mill, J.S.** (1843) *System of Logic: Ratiocinative and Inductive*, in *Collected Works of John Stuart Mill*, vols 7 and 8, London: Routledge, 1991.(Sometimes said to be the most-published logic book ever perhaps since surpassed by Copi's *Introduction to Logic*); an accessible, well-organized work often lamented for its lack of rigour, poor sense of history, and undeserved influence.)
- Peano, G.** (1973) *Selected Works of Giuseppe Peano*, trans. and ed. H. Kennedy, Toronto, Ont.: University of Toronto Press.(A good selection in English of the works of Peano, whose notation and conception of logic were probably more influential than any other nineteenth-century logician; a very small collection.)
- Peirce, C.S.** (1867-90) 'On an Improvement in Boole's Calculus of Logic' and other important logical papers, in *Collected Papers*, vol. 3, ed. C. Hartshorne, P. Weiss and A.W. Burks, Cambridge, MA: Harvard University Press, 1933; and in *Writings of Charles S. Peirce: A Chronological Edition*, ed. M. Fisch et al., Bloomington, IN: Indiana University Press, 1982-.(The *Collected Writings* includes the more carefully edited versions - at

least of works up to 1886.)

- Schröder, E.** (1877) *Der Operationskreis des Logikkalküls*, Leipzig: Teubner. (Schröder's first, brief and elegant work on logic; shows familiarity with works by Boole and Grassmann, but not yet with works by Peirce or De Morgan, or with the theory of relations.)
- Schröder, E.** (1890-1905) *Vorlesungen über die Algebra der Logik (Lectures on the Algebra of Logic)*, Leipzig: Teubner, 3 vols; repr. New York: Chelsea, 1966. (The monumental *locus classicus* of the algebra of logic, including relations; a highly systematic work that totals thousands of pages and hundreds of theorems; unindexed, it remains untranslated and has probably been read in its entirety by only a handful of German-speaking readers. See Dipert (1990-1) for discussion.)
- Styazhkin, N.I.** (1964) trans. *History of Mathematical Logic from Leibniz to Peano*, Cambridge, MA, and London: MIT Press, 1969. (A briefer treatment of nineteenth-century logic than Kneale and Kneale (1962), but often more insightful and knowledgeable in its discussions of symbolisms, of German logic in the eighteenth and nineteenth centuries, and of Russian figures such as Poretskii.)
- Trendelenburg, F.A.** (1847-67) *Historische Beiträge zur Philosophie*, Berlin: Bethge. (Together with his *Logische Untersuchungen* and editions of Aristotle's *Prior Analytics*, this book began interest in the logics of Leibniz and Aristotle in the German-speaking world; probably read in part by Frege.)
- Venn, J.** (1881) *Symbolic Logic*, London: Macmillan, 2nd edn, 1894; repr. New York: Chelsea, 1979. (By the time it appeared, this work was unoriginal, but usefully summarized Boolean logic in a book-length form more readable than Boole's own works; the first Boolean 'textbook'.)
- Whately, R.** (1826) *Elements of Logic*, Delmar, NY: Scholars' Facsimiles and Reprints, 1975. (A highly intelligent, succinct, non-symbolic work that represents the high-water mark of the non-symbolic English 'logic textbook' tradition.)

Logic in the early 20th century

The creation of modern logic is one of the most stunning achievements of mathematics and philosophy in the twentieth century. Modern logic - sometimes called logistic, symbolic logic or mathematical logic - makes essential use of artificial symbolic languages.

Since Aristotle, logic has been a part of philosophy. Around 1850 the mathematician Boole began the modern development of symbolic logic. During the twentieth century, logic continued in philosophy departments, but it began to be seriously investigated and taught in mathematics departments as well. The most important examples of the latter were, from 1905 on, Hilbert at Göttingen and then, during the 1920s, Church at Princeton.

As the twentieth century began, there were several distinct logical traditions. Besides Aristotelian logic, there was an active tradition in algebraic logic initiated by Boole in the UK and continued by C.S. Peirce in the USA and Schröder in Germany. In Italy, Peano began in the Boolean tradition, but soon aimed higher: to express all major mathematical theorems in his symbolic logic. Finally, from 1879 to 1903, Frege consciously deviated from the Boolean tradition by creating a logic strong enough to construct the natural and real numbers. The Boole-Schröder tradition culminated in the work of Löwenheim (1915) and Skolem (1920) on the existence of a countable model for any first-order axiom system having a model.

*Meanwhile, in 1900, Russell was strongly influenced by Peano's logical symbolism. Russell used this as the basis for his own logic of relations, which led to his logicism: pure mathematics is a part of logic. But his discovery of Russell's paradox in 1901 required him to build a new basis for logic. This culminated in his masterwork, *Principia Mathematica*, written with Whitehead, which offered the theory of types as a solution.*

Hilbert came to logic from geometry, where models were used to prove consistency and independence results. He brought a strong concern with the axiomatic method and a rejection of the metaphysical goal of determining what numbers 'really' are. In his view, any objects that satisfied the axioms for numbers were numbers. He rejected the genetic method, favoured by Frege and Russell, which emphasized constructing numbers rather than giving axioms for them. In his 1917 lectures Hilbert was the first to introduce first-order logic as an explicit subsystem of all of logic (which, for him, was the theory of types) without the infinitely long formulas found in Löwenheim. In 1923 Skolem, directly influenced by Löwenheim, also abandoned those formulas, and argued that first-order logic is all of logic.

Influenced by Hilbert and Ackermann (1928), Gödel proved the completeness theorem for first-order logic (1929) as well as incompleteness theorems for arithmetic in first-order and higher-order logics (1931). These results were the true beginning of modern logic.

1 What is modern logic?

Modern logic crystallized gradually during the period 1850-1950. It was not created by any one person or school, but evolved from the work of several. We now recognize that the following features were needed for its mature development:

- (a) There is a formal language whose symbols separate it from ordinary language.
- (b) In the formal language, L , the syntax is clearly separated from the semantics. The syntax is like the grammar of a natural language, specifying which sequences of symbols (called 'formulas') are permissible. Meaning does not occur in the syntax but in the semantics. Thus logical concepts such as sentence, axiom, formal proof and consistency are part of the syntax. Concepts such as truth, satisfiability and definability are part of the semantics. Formal proofs can be checked effectively and do not depend on the meaning of the terms in them.
- (c) Within the syntax of L , the logical symbols are clearly separated from the non-logical symbols. The logical symbols include connectives such as 'and', 'not', 'if...then...'. The meaning of the non-logical symbols changes with different interpretations, while that of the logical symbols does not. The non-logical symbols may include individual constants, which, in the semantics, are interpreted as specific individuals. There may also be function constants and relation constants, which are interpreted as specific functions and relations. Besides the logical and non-logical symbols, there are variables, which may be free or bound.

- (d) The syntax gives a recursive definition for the formulas of L .
- (e) There is a clear distinction between the object language L , in which our formulas occur, and the metalanguage, in which we speak about L . Within the metalanguage, we must define what truth means in L . Theorems in the metalanguage are called metatheorems.
- (f) Usually there will be levels within our symbolic logic. The lowest level (called propositional logic) is that of the connectives ‘and’, ‘not’, and so on. The next level will have individual variables x and quantifiers (‘for all x ’, ‘there exists an x ’) over those variables. If that is the highest level, then we have first-order logic. But there may also be function variables (or relation variables) and quantifiers over them, giving second-order logic (see **Second- and higher-order logics**). In the simple theory of types, there are individual variables, function variables, variables for functions of functions, and so on, and this is sometimes called ω -order logic. We must decide which order our logic will have.
- (g) We must also specify other matters. Will the semantics of our logic have only two truth-values, ‘true’ and ‘false’, as in traditional logic, or will it have three or more truth-values (see **Many-valued logics**)? Will our logic be ‘classical’ (in which the law of the excluded middle holds) or ‘intuitionistic’ (in which that law fails; see **Intuitionism**)? Will our logic be finitary (like all the logics discussed above) or will it be infinitary (that is, have infinitely long formulas, or have variable-binding operators that give the same effect, or have rules of inference with infinitely many premises; see **Infinitary logics**)?
- (h) Finally, we must study the metatheory of L , that is, what general results can be proved, in the metalanguage, to be true of L . This includes the consistency and independence of the logical axioms for L as well as whether L is ‘complete’ (that is, whether every true sentence of L is provable in L).

2 Logic in 1900

In 1900 there were five main approaches to logic. The first of these was traditional Aristotelian logic, in which the syllogism was central. It had split into several branches, encompassing the idealist logic of Bradley and the pragmatist logic of Dewey. The remaining approaches differed from the first by using symbols in a fundamental way.

The second approach was that of Boole. In 1854 Boole had introduced a logical calculus, whose symbols were borrowed from mathematics, and given his calculus three interpretations: propositions, classes and probability. Around 1900, Boole’s active successors were Peirce and Schröder. During the 1880s Peirce had modified Boole’s logic by introducing quantifiers and relations. The second approach was regarded as trivial by certain Oxford logicians who espoused the traditional one (see Passmore 1968: 156, 240).

The third approach originated in Italy. There Peano began his work on logic (1888) within the Boolean tradition, influenced by Schröder (1877). To avoid confusing Schröder’s logical symbols ‘ $\cdot$ ’ and ‘ $+$ ’ (for ‘and’ and ‘or’, respectively) with the corresponding arithmetical symbols for addition and multiplication, Peano replaced them with his own, derived from Hermann Grassmann: ‘ $\cap$ ’ and ‘ $\cup$ ’. Peano retained a Boolean ambiguity by interpreting ‘ $\cap$ ’ and ‘ $\cup$ ’ not only as ‘and’ and ‘or’ between propositions but also as ‘intersection’ and ‘union’ of classes. (In 1906 Russell removed this ambiguity by keeping ‘ $\cap$ ’ and ‘ $\cup$ ’ for ‘intersection’ and ‘union’ while introducing a new symbol ‘ $\vee$ ’ for ‘or’ and using a dot for ‘and’; nevertheless, like Frege, he had no non-logical symbols and so did not satisfy requirement (c) above.) Peano went beyond the Boolean tradition by symbolizing much of mathematics.

The fourth approach, that of Frege, was isolated from the others. Frege proposed a second-order symbolic language for treating mathematics - particularly the real and complex numbers - in his *Begriffsschrift* (1879). There he replaced the traditional logical notions of subject and predicate by the more mathematical ones of function and argument. In the culmination of his work (*Grundgesetze der Arithmetik*, 1893 and 1903), he attempted to establish that all truths of arithmetic (but not geometry) can be proved from logic alone (see **Logicism**). To do so, he used a very strong logic, which in 1902 turned out to be inconsistent. No doubt his daunting two-dimensional symbolism partly accounts for his limited influence at the time. But Frege was also isolated by his explicit refusal to work in the Boolean tradition, as well as by his many innovations and his severe criticisms of other logicians.

The fifth approach to logic, that of Hilbert, was rooted in the foundations of geometry. In 1899 he gave axioms for Euclidean geometry that were purely formal, although not yet expressed in symbols. He proposed three main tasks for such an axiom system: establishing its consistency, independence and completeness. (At that time, ‘completeness’ was imprecise and required only that the axioms imply all known theorems of geometry.)

So at the turn of the century there was no unified tradition in logic and no common view of what logic is and what its goals are; at the same time, it was generally agreed that there is only one correct logic. This fragmented situation continued until at least 1930 (although the Boole-Schröder tradition was strong until the 1920s), when the theory of types became increasingly important. After 1930, first-order logic gradually replaced the theory of types.

3 The algebraic tradition of Boole and Schröder

During the 1890s Schröder continued the Boolean tradition (see §2 above). Schröder’s logic began with his ‘identical calculus’ or lattice theory (1890: 161); it was related to a calculus of classes, one of domains and one of propositions. In 1895, his calculus of relations elaborated that of Peirce.

Although a critic of Frege, Schröder too regarded arithmetic as a part of deductive logic (1890: v) and later went beyond Frege’s version of logicism by stating that ‘I consider pure Mathematics to be only one branch of general Logic’ (1898: 46).

The Boole-Schröder tradition culminated in the work of Löwenheim and Skolem. Löwenheim’s most important paper (1915), on Schröder’s calculus of relations, distinguished between first-order and second-order logic while retaining elements of infinitary logic (see §10 below). Löwenheim’s semantic theorem was that if a first-order formula is true in all countable domains, then it is true in all domains. Löwenheim was well aware that this result failed in his full logic, which was a kind of infinitary second-order logic.

In 1920 Skolem extended Löwenheim’s theorem, and in this form it is known as the ‘Löwenheim-Skolem theorem’: if a countable set of first-order formulas has a model, it has a countable model. This was the first major metatheorem for first-order logic.

4 Peano’s tradition

The most important of Peano’s logical accomplishments was to invent a clear symbolism. Much of it - such as $\in$ for class membership and $\supset$ for material implication - continues today, thanks to its adoption by Russell. By contrast, the logical symbolisms of Frege and Schröder are all but extinct.

In 1891 Peano believed that he could express any logical relation by using seven primitive symbols: $\in$ (‘is’), $=$ (‘is equal’), $\supset$ (‘implies’), $\cap$ (‘and’), $\cup$ (‘or’), $-$ (‘not’) and $\wedge$ (‘absurd’). He was mistaken, as Frege remarked. Partly under Frege’s influence, Peano used nine kinds of primitive symbol in 1897: the old signs $\in$, $\supset$, $\cap$, $-$, together with the signs K (‘class’), (x, y) (‘ordered pair’), $=_{\text{Def}}$ (‘definition’), parentheses and variables. Yet both Frege and Peano accepted the principle of comprehension (that is, every predicate determines the class of those things satisfying the predicate), which eventually led to the paradoxes.

Peano tried to render mathematics completely unambiguous by expressing it in his symbolism. He first used his symbolism in geometry (1888), next in his postulates for the natural numbers (1889), and then he applied it to differential equations (1890). He assembled a group of Italian collaborators to create the encyclopedic *Formulaire de mathématiques* (1894-1908), which aimed to symbolize all major mathematical theorems.

Among those collaborators was Alessandro Padoa, who investigated the undefinability of a term from other terms in a given axiom system (1901). Padoa emphasized that the notion ‘definable’ was not absolute but depended on the particular axiom system and its primitive symbols. This insight was overlooked when the paradoxes of definability, such as Richard’s paradox, were discovered soon afterward (see [Paradoxes of set and property §6](#)).

5 Russell’s response to Peano and Frege

In 1900, before he discovered the paradoxes, Russell adopted Peano’s symbolism and used it to develop his own logic of relations. Russell was particularly impressed by Peano’s distinction between the membership relation $\in$ and the relation $\supset$ construed as inclusion. Peano did not distinguish, however, between the implication $p \supset q$ and the inference ‘From the proposition p infer the proposition q ’, although this distinction had already been made by

Frege (1879). The deduction theorem - a metatheorem showing the correct relation between the two - was only proved by Tarski in about 1921.

For Russell, symbolic logic had three parts: the calculi of propositions, of classes and of relations. In the *Principles of Mathematics*, he gave primacy to the propositional calculus over the calculus of classes (1903: 12), and also proposed his logicism, which went beyond Frege's by including geometry and parts of mechanics: all mathematical concepts are definable from the basic concepts of logic, and all mathematical theorems are provable from the axioms of logic.

In 1903, after finishing the *Principles*, Russell shifted from an analysis of logic in the spirit of Peano to one in the spirit of Frege. Peano had given the universal quantifier only in the form $\phi x \supset_x \psi x$, that is, 'For every x , ϕx implies ψx '. Russell introduced a new notation for an idea that he took from Frege: $(x)\phi x$, that is, 'For every x , ϕx '. Likewise, Peano had been able to express an existential quantifier only in the form $\exists a$, that is, 'The class a is non-empty'. Russell introduced a Fregean existential quantifier: $(\exists x)\phi x$, that is, 'There is an x such that ϕx '. Russell's notation later became standard. It had the virtue of disentangling quantifiers from the quantifier-free part of a formula and of disentangling logical formulas from the notion of class. Finally, Russell adopted Frege's assertion sign, ' $\vdash$ '. Thus, for Russell, $\vdash (x)\phi x$ meant 'It is true that $(x)\phi x$ '. (Nowadays $\vdash$ does not mean 'it is true that' but rather, 'it is provable that', and indicates a theorem of logic; see Church (1956: 82).)

6 Hilbert, the geometric tradition and the American postulate theorists

Hilbert's interest in logic stemmed from his devotion to the axiomatic method, which he first exploited to give a definitive form to the axioms of Euclidean geometry in 1899. Here he continued the geometric tradition by using models to establish the consistency and independence of his axioms. Yet he distanced himself from the traditional philosophical concern with what points or numbers 'really' are by treating them as unspecified 'things', defined only by the axioms. As long as the 'things' satisfied the axioms, he insisted, it did not matter if they were points, lines and planes, or tables, chairs and mugs. In accordance with requirement (b) in §1 above, the syntax was separated clearly from the semantics.

At that time Hilbert corresponded with Frege on the foundations of geometry. Frege held the traditional view that the geometric axioms express intuitions, while Hilbert supported the modern view that an axiom system at most determines the 'things' mentioned up to isomorphism (see Nagel 1939). Frege insisted that the only way to prove the consistency of an axiom system was to give a model, while Hilbert believed that the consistency of a system yields the existence of a model in which it is 'true' (see Frege 1980: 39). Brouwer disputed Hilbert's claim in 1929. Around 1930, Gödel proved this claim for first-order logic, but, using incompleteness, refuted it for second-order logic (see §9).

From 1900 until 1940, the axiomatic method, as formulated by Hilbert and Peano, was used by a group of mathematicians sometimes called the American postulate theorists (see Scanlan 1991) to axiomatize a variety of mathematical systems - including the complex numbers, Boolean algebra, and fields - and to prove the consistency and independence of the axioms. This group included Huntington, Veblen and E.H. Moore. In 1902, while giving an axiomatization for the real numbers, Huntington isolated the idea that this system has, up to isomorphism, only one model. Such a system is 'categorical' (to use the name that Veblen gave it in 1904). A year later, Huntington conjectured that every categorical system S is deductively complete, that is, every sentence expressible in S is provable or refutable in S . It was not until Gödel's incompleteness theorems that Huntington was proved wrong. Moreover, in first-order logic, any system having an infinite model is not categorical (see §8).

7 *Principia Mathematica*

Between 1902 and 1913 Russell and Whitehead wrote *Principia Mathematica* (1910-13), which aimed to give strong evidence for logicism by axiomatizing logic (via the theory of types) and deducing much of pure mathematics from it. In doing so, Russell was particularly concerned to avoid the paradoxes of logic. He also shared with Peano a desire to minimize the number of axioms and primitive concepts.

Principia Mathematica strongly influenced the development of logic, but most of that influence did not come until the 1920s. When its second edition appeared in 1925, Russell considered the most important improvement to logic to be the reduction by Sheffer of the primitive concepts of propositional logic from two ('or' and 'not') to one

(‘neither... nor’) and the reduction by Nicod of its axioms from five to one. (Ironically, later logical research hardly used such reductions.) But in that edition Russell did not mention a more important advance made by Wiener (1914) and known to Russell: the definition of ordered pair (and hence of relation) in terms of classes.

Not everyone accepted the theory of types. Veblen argued that ‘formal logic has to be taken over by mathematicians.... There does not exist an adequate logic at the present time, and unless the mathematicians create one, no one else is likely to do so’ (1925: 141).

Already in the *Principles* (1903: §17) Russell believed it impossible to establish the independence of one axiom of logic from the others. In *Principia Mathematica*, Whitehead and Russell retained this conviction (1910: 95). For Russell, as for Frege, one cannot stand outside of logic, and so it is not surprising that *Principia Mathematica* lacked a metalanguage. This notion of metalanguage originated with Hilbert (1922), from whom it was adopted by Tarski (1930) and Carnap (1934). It was natural when a system of logic was viewed as simply another mathematical system. (Russell (1922) independently considered levels of language.) Nowadays, the most important theorems of logic are metatheorems, that is, theorems in the metalanguage about the system of logic, and they involve the sort of notions that Hilbert used in his theorems about geometry: consistency, independence and completeness.

8 The emergence of first-order logic

By 1955, first-order logic was dominant among logicians who were mathematicians (though the theory of types remained important among philosophers). However, first-order logic was slow in coming. It first clearly emerged as a ‘subsystem’ of logic in Hilbert’s lectures of 1917, published in Hilbert and Ackermann (1928). (Löwenheim had separated a subsystem of logic from his full system in 1915, but his subsystem allowed quantifiers to be expanded into infinitely long formulas, and so was richer than first-order logic, in which all formulas have finite length.)

Hilbert, unlike Löwenheim, was very careful to specify his syntax, and gave, for the first time, a recursive definition of first-order formula (see (d) in §1 above). Hilbert also stated axioms for first-order logic. While he had borrowed his axioms for propositional logic from *Principia Mathematica*, first-order logic was his own. Moreover, he differed from Russell by posing the metamathematical questions of the consistency, independence and completeness of his axioms. He proved the consistency of both propositional and first-order logic, but was unable to answer the other questions. Hilbert wanted to give axiomatizations for various theories within logic, whereas Russell wanted a strong logic in which to deduce those theories from his logical axioms. Hilbert’s long-term goal was to prove the consistency of all of classical mathematics by finitary means (see *Hilbert’s programme and formalism*.)

In 1917 Hilbert persuaded Bernays to work with him on logic. Bernays’ first results (1918, published only in 1926) established the independence and completeness of propositional logic. Meanwhile, Post had independently found and published them (1921).

Also in 1917, Hilbert emphasized the ‘*Entscheidungsproblem*’: to find a decision procedure for every mathematical problem. Within the Schröderian tradition, decision procedures had already been studied by Löwenheim (1915), who solved the *Entscheidungsproblem* for monadic predicates within first-order logic, a solution improved by Skolem (1919). Within Hilbert’s school, Bernays gave a decision procedure for propositional logic in 1918, and in 1928 argued that the *Entscheidungsproblem* was the central problem of logic (see *Proof theory §2*). Work on the *Entscheidungsproblem* consisted of finding a decision procedure for more and more complicated classes of first-order sentences in prenex normal form (that is, consisting of a quantifier-free matrix preceded by a string of quantifiers), where the complexity of the class was measured by the number of quantifiers in the prefix.

Meanwhile, in 1923, Skolem abandoned the infinitely long formulas adopted from Schröder and Löwenheim, and began working explicitly in first-order logic. Skolem made the radical proposal that the membership relation ‘ $\in$ ’ of set theory be treated, not as a part of logic in the way that Peano and Russell had done, but like any other relation (that is, as a non-logical symbol). In logic, such relations could be given various interpretations, and so should ‘ $\in$ ’. This proposal resulted in the so-called ‘Skolem paradox’ of first-order logic: set theory has a countable model although it contains uncountable sets (see *Paradoxes of set and property §9*).

In France, Herbrand (1930) made a major contribution to the relation between propositional logic and first-order logic.

9 The definitive metatheorems: Gödel on completeness and incompleteness

The most important person influenced by Hilbert's logic was Gödel, whose doctoral dissertation (1929) solved Hilbert and Ackermann's problem of showing, for countable languages, the completeness of first-order logic: every valid first-order formula is provable. In fact, Gödel proved the stronger result that every consistent set of first-order formulas has a model (1929).

When Gödel's completeness theorem was published in 1930, he included as a corollary the 'compactness theorem': a set of first-order formulas has a model if every finite subset has a model. The importance of this theorem was only appreciated later by Maltsev (1936) and Henkin (1947) when they introduced uncountable languages.

Even more important were Gödel's incompleteness theorems (1931). His first incompleteness theorem stated that if a 'decidable' set Σ of formulas implies the axioms for the natural numbers with addition and multiplication, then there is a sentence in the language of Σ that is neither provable nor refutable. (Here 'decidable' means that there is a mechanical procedure for determining whether a formula is in Σ .) The second incompleteness theorem stated that such an axiom system Σ cannot prove its own consistency (see [Gödel's theorems](#)).

The incompleteness theorems were devastating. They refuted Russell's logicism by showing that there are true mathematical sentences which cannot be proved in the theory of types. Likewise, they refuted Hilbert's programme, since the consistency of number theory cannot be established except by a stronger theory. Church soon established that there is no decision procedure for first-order logic. Finally, the incompleteness theorems led to attempts, by Carnap and Zermelo, to circumvent them by introducing an infinitary logic (see §10).

10 The prehistory of infinitary logic

Peirce and Schröder treated quantifiers in a way that differs from the modern one. If we fix a finite set with n elements $a_1, a_2, \dots, a_n$ as our universe D of discourse and let Px mean that x has the property P , then in D we can express 'All x have property P ' by ' Pa_1 and Pa_2 and... and Pa_n '. If D is infinite and its elements are $a_1, a_2, \dots, a_n, \dots$, then Peirce and Schröder still treat 'All x have property P ' as equivalent to ' Pa_1 and Pa_2 and... and Pa_n and...'. (An analogous statement holds in D for 'Some x has P ' and ' Pa_1 or Pa_2 or... or Pa_n or...'.) Thus a quantifier could be expanded as an infinite conjunction or disjunction. This was the beginning of an infinitary logic, but one that was not yet formulated as a distinct kind of logic.

Here we see two distinct ways of treating quantifiers: (1) 'For all x , Px ' and (2) the infinitely long conjunction ' Pa_1 and Pa_2 and... and Pa_n and...'. Frege, Peano and Russell used quantifiers only in the first way. By contrast, Peirce and Schröder used both ways interchangeably. Hilbert (1905) was influenced by Schröder to define quantifiers in the second way when the universe D was the natural numbers. In his mature work (1923), Hilbert still viewed quantifiers as going beyond the finitary.

Meanwhile, within the Schröderian tradition, Löwenheim used not only such infinitely long conjunctions and disjunctions but also infinite strings of quantifiers, in his paper proving Löwenheim's theorem. Skolem continued to use this infinitary logic when in 1920 he extended Löwenheim's theorem to a countable set of first-order formulas, but by 1923 he abandoned infinitary logic in favour of first-order logic.

Infinitary logic arose, independently, in Zermelo's paper (1931) reacting against first-order logic and specifically against Skolem's argument that there is a countable model for set theory (see [Paradoxes of set and property §9](#)). Zermelo formulated a strong infinitary logic in which, he hoped, the Löwenheim-Skolem theorem was false. Soon he had a second reason to propose an infinitary logic, namely to avoid Gödel's incompleteness theorems and to have every valid sentence be provable. But Zermelo's logic was adopted by no one.

Hilbert, Tarski and Zermelo each formulated a different version of an infinitary rule of inference, now known as the ω -rule. (Influenced by Hilbert and Tarski, Carnap also used the ω -rule.) This rule requires the language of logic to have infinitely many names, one for each natural number n . If a formula $\phi(a_n)$ has been proved for the

name a_n of each natural number n , the ω -rule infers ‘For every x , $\phi(x)$ ’. Without the ω -rule, an axiomatization of the natural numbers could be ω -inconsistent, that is, prove $\phi(a_n)$ for each natural number n but also prove ‘There is some x such that not $\phi(x)$ ’. Unlike Carnap, Hilbert and Zermelo, all of whom supported this rule, Tarski was occasionally sceptical of it.

Only in 1955 were infinitely long formulas accepted as a legitimate part of mathematical logic - thanks to Henkin, Tarski and Karp - in apparent ignorance of earlier work on such formulas by Zermelo and others. By this time, first-order logic was firmly established as the logical framework for mathematics, and those infinitary formulas were considered within various extensions of first-order logic. A different kind of extension was proposed by Mostowski (1957): generalized quantifiers. One such quantifier was ‘there exist uncountably many’. For that quantifier, Mostowski asked if it is possible to prove a completeness theorem analogous to that for first-order logic. Later the answer turned out to be ‘yes’.

11 A glance past Gödel: recursion theory, model theory, set theory

Modern logic began with Gödel’s completeness and incompleteness theorems. Today they are the goal of most introductory logic courses. From his incompleteness theorems, recursion theory emerged during the 1930s, when the notion of ‘computable’ was made precise by the Gödel-Herbrand notion of recursive function and, equivalently, by Turing machines (see [Computability theory](#)).

Model theory emerged as a distinct part of logic around 1950, although its first important result (Löwenheim’s theorem) dates back to 1915. One fundamental notion - relational structure - was already vaguely present in Hilbert (1899). A relational structure is an ordered pair consisting of a set and a family of relations on that set; such structures include groups, fields and Boolean algebras, but not topologies. Model theory began to develop into its modern form through the discovery that the strong completeness theorem for first-order logic (that is, any consistent set of first-order sentences has a model) implies results about structures. Connected with this was Henkin’s use of languages with uncountably many symbols (1947) and Tarski’s classification (1952) of structures up to ‘elementary equivalence’ (that is, being indistinguishable in first-order logic).

Meanwhile, set theory developed much closer ties with logic. In 1938 Gödel used first-order logic to prove the important result that if the usual axioms of set theory are consistent, they remain so when the axiom of choice and the generalized continuum hypothesis are assumed (see [Axiom of choice](#); [Set theory](#)). Gödel’s idea was to consider, not all sets, but only those sets first-order definable from sets of previous levels (see [Constructible universe](#)). And in 1963 first-order logic was essential to Cohen’s proof that the axiom of choice and the generalized continuum hypothesis are independent of the usual axioms of set theory (see [Forcing](#)).

12 Coda: what is modern logic?

Until about 1850, it was fairly easy to say what logic was. Today it is much harder, although classical first-order logic remains dominant. After 1920, many-valued logics (that is, those having three or more truth-values) were investigated by Post and Łukasiewicz. Likewise, modal logic was seriously studied, thanks to C.I. Lewis (1918; see [Modal logic](#)). And during the 1950s infinitary logic crystallized. Finally, there are now many specialized logics (see, for example, [Quantum logic](#)).

During the 1970s first-order logic was characterized by ‘Lindström’s theorem’: first-order logic is the strongest logic that contains propositional logic and for which the compactness theorem and the Löwenheim-Skolem theorem hold. Thus any logic which extends first-order logic must violate the compactness theorem or the Löwenheim-Skolem theorem. An idea of the wide range of modern logic can be obtained by consulting Barwise (1977), Barwise and Feferman (1985), and Gabbay and Guenther (1983-9).

See also: [Logical and mathematical terms, glossary of](#)

GREGORY H. MOORE

References and further reading

- Barwise, J.** (ed.) (1977) *Handbook of Mathematical Logic*, Amsterdam: North Holland. (An excellent survey of modern logic.)
- Barwise, J. and Feferman, S.** (eds) (1985) *Model-Theoretic Logics*, New York: Springer. (The best survey of

infinitary logic.)

- Bernays, P.** (1918) 'Beiträge zur axiomatischen Behandlung des Logik-Kalküls' (Contributions to the Axiomatic Treatment of the Logical Calculus), unpublished *Habilitationsschrift*, Göttingen University; published in much revised form, 'Axiomatische Untersuchung des Aussagen-Kalküls der *Principia Mathematica*' (Axiomatic Investigation of the Propositional Calculus of *Principia Mathematica*), *Mathematische Zeitschrift* 25: 305- 20, 1926.(First proof of the completeness of propositional logic; compare with Post (1921).)
- Carnap, R.** (1934) *Logische Syntax der Sprache*, Vienna: Springer; revised and trans. A. Smeaton, *The Logical Syntax of Language*, London: Kegan Paul, and New York: Harcourt Brace, 1937.(A precise version of the theory of types and the ω -rule.)
- Church, A.** (1956) *Introduction to Mathematical Logic*, Princeton, NJ: Princeton University Press.(An erudite but useful introduction to modern logic.)
- Frege, G.** (1879) *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*, Halle: Nebert; trans. 'Begriffsschrift, a Formula Language, Modelled upon That of Arithmetic, for Pure Thought', in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967, 1-82.(Frege's first, ground-breaking work.)
- Frege, G.** (1893, 1903) *Grundgesetze der Arithmetik: begriffsschriftlich abgeleitet*, Jena: Pohle, 2 vols; repr. as *Grundgesetze der Arithmetik*, Hildesheim: Olms, 1966; Part 1 of vol. 1 trans. M. Furth, *Basic Laws of Arithmetic: An Exposition of the System*, Berkeley, CA: University of California Press, 1964; extracts from vol. 2, including 'Frege on Russell's Paradox', in *Translations from the Philosophical Writings of Gottlob Frege*, trans. and ed. P.T. Geach and M. Black, Oxford: Blackwell, 3rd edn, 1980. (Volume 1 gives a detailed development of his system of logic. Volume 2 has Frege's reaction to Russell's paradox.)
- Frege, G.** (1980) *Philosophical and Mathematical Correspondence*, trans. H. Kaal, ed. G. Gabriel *et al*, Chicago, IL: University of Chicago Press.(Translation of Frege's letters. The correspondence with Hilbert and Russell is especially interesting.)
- Gabbay, D. and Guenther, F.** (eds) (1983-9) *Handbook of Philosophical Logic*, Dordrecht: Reidel, 4 vols.(A standard modern source on classical and non-classical logics.)
- Gödel, K.** (1929) 'Über die Vollständigkeit des Logikkalküls', doctoral dissertation, University of Vienna; trans. 'On the Completeness of the Logical Calculus', in *Collected Works*, vol. 1, *Publications 1929-1936*, ed. S. Feferman, J.W. Dawson Jr, S.C. Kleene, G.H. Moore, R.M. Solovay, and J. van Heijenoort, New York and Oxford: Oxford University Press, 1986, 44-101.(First proof of the completeness theorem for first-order logic.)
- Gödel, K.** (1930) 'Die Vollständigkeit der Axiome des logischen Funktionenkalküls', *Monatshefte für Mathematik und Physik* 37: 349-60; trans. 'The Completeness of the Axioms of the Functional Calculus of Logic', in *Collected Works*, vol. 1, *Publications 1929-1936*, ed. S. Feferman, J.W. Dawson Jr, S.C. Kleene, G.H. Moore, R.M. Solovay, and J. van Heijenoort, New York and Oxford: Oxford University Press, 1986, 102-23; and in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967, 582-91.(First published proof of the completeness of first-order logic, together with the compactness theorem.)
- Gödel, K.** (1931) 'Über formal unentscheidbare Sätze der *Principia Mathematica* und verwandter Systeme I', *Monatshefte für Mathematik und Physik* 38: 173-98; trans. 'On Formally Undecidable Propositions of *Principia Mathematica* and Related Systems', in *Collected Works*, vol. 1, *Publications 1929-1936*, ed. S. Feferman, J.W. Dawson Jr, S.C. Kleene, G.H. Moore, R.M. Solovay, and J. van Heijenoort, New York and Oxford: Oxford University Press, 1986, 126-95; and in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967, 592-617.(First proof of Gödel's incompleteness theorems.)
- Grattan-Guinness, I.** (1981) 'On the Development of Logics between the Two World Wars', *American Mathematical Monthly* 88: 495-509.(Overview of logic from 1920 to 1940.)
- Grattan-Guinness, I.** (1985) 'Russell's Logicism versus Oxbridge Logics, 1890-1925: A Contribution to the Real History', *Russell: The Journal of the Bertrand Russell Archives*, new series, 5: 101-31.(Traditional logic versus the symbolic logic of Peano and Russell.)
- Heijenoort, J. van** (ed.) (1967) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press.(Includes translations of many relevant articles.)
- Henkin, L.** (1947) 'The Completeness of Formal Systems', unpublished doctoral dissertation, Princeton University.(First proof of the completeness theorem for uncountable languages.)
- Herbrand, J.** (1930) 'Recherches sur la théorie de la démonstration', *Prace Towarzystwa Naukowego Warszawskiego*, Wydział 3, no. 33; trans. 'Researches on Proof Theory', in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967,

- 525-81. (Herbrand's theorem.)
- Hilbert, D.** (1899) *Grundlagen der Geometrie*, Leipzig: Teubner, 7th edn, 1930; 2nd edn trans. L. Unger and P. Bernays, *Foundations of Geometry*, La Salle, IL: Open Court, 1971. (Consistency and independence results in geometry; the modern axiomatic method.)
- Hilbert, D.** (1905) 'Über die Grundlagen der Logik und der Arithmetik', in A. Krazer (ed.) *Verhandlungen des dritten internationalen Mathematiker-Kongresses in Heidelberg vom 8. bis 13. August 1904*, Leipzig: Teubner, 174-85; trans. 'On the Foundations of Logic and Arithmetic', in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967, 129-38. (Hilbert's first attempt to prove the consistency of arithmetic.)
- Hilbert, D.** (1922) 'Neubegründung der Mathematik. Erste Mitteilung' (New Foundations of Mathematics. First Communication), *Abhandlungen aus dem mathematischen Seminar der Hamburgischen Universität* 1: 157-77; trans. 'The New Grounding of Mathematics', in W.B. Ewald (ed.) *From Kant to Hilbert: A Source Book in the Foundations of Mathematics*, vol. 2, Oxford: Clarendon Press, 1996, 1115-34. (His first explicit use of 'metamathematics'.)
- Hilbert, D.** (1923) 'Die logischen Grundlagen der Mathematik' (The Logical Foundations of Mathematics), *Mathematische Annalen* 88: 151-65; trans. 'The Logical Foundations of Mathematics', in W.B. Ewald (ed.) *From Kant to Hilbert: A Source Book in the Foundations of Mathematics*, vol. 2, Oxford: Clarendon Press, 1996, 1134-48. (His mature work on proof theory.)
- Hilbert, D. and Ackermann, W.** (1928) *Grundzüge der theoretischen Logik*, Berlin: Springer, 2nd edn, 1938; trans. L.M. Hammond, G.G. Leckie and F. Steinhardt, *Principles of Mathematical Logic*, ed. R.E. Luce, New York: Chelsea, 1950. (First modern textbook in mathematical logic.)
- Lewis, C.I.** (1918) *A Survey of Symbolic Logic*, Berkeley, CA: University of California Press; revised edn, New York: Dover, 1960. (A historical account of modern logic, together with an axiomatization of modal logic.)
- Löwenheim, L.** (1915) 'Über Möglichkeiten im Relativkalkül', *Mathematische Annalen* 76: 447-70; trans. S. Bauer-Mengelberg, 'On Possibilities in the Calculus of Relatives', in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967, 232-51. (First important results on the semantics of first-order and infinitary logic.)
- Maltsev, A.I.** (1936) 'Untersuchungen aus dem Gebiete der mathematischen Logik', *Matematicheskii Sbornik*, new series, 1: 323-36; trans. 'Investigations in the Area of Mathematical Logic', in B. Wells (ed.) *The Metamathematics of Algebraic Systems: Collected Papers, 1936-1967*, Amsterdam: North Holland, 1971. (First use of the compactness theorem for uncountable first-order languages.)
- Moore, G.H.** (1980) 'Beyond First-Order Logic: The Historical Interplay Between Mathematical Logic and Axiomatic Set Theory', *History and Philosophy of Logic* 1: 95-137. (The influence of logic on set theory, and vice versa.)
- Moore, G.H.** (1988) 'The Emergence of First-Order Logic', in W. Aspray and P. Kitcher (eds) *History and Philosophy of Modern Mathematics*, Minneapolis, MN: University of Minnesota Press, 95-135. (How first-order logic emerged from richer logics, including higher-order and infinitary logics.)
- Moore, G.H.** (1997) 'The Prehistory of Infinitary Logic: 1885-1955', in *Tenth International Congress of Logic, Methodology, and Philosophy of Science, Florence, August 1995*, vol. 2, M.L. Dalla Chiara, K. Doets, D. Mundici and J. van Benthem (eds) *Structures and Norms in Science*, Dordrecht: Kluwer, 105-23. (The early history of infinitary logics from Peirce through Löwenheim and Skolem to Novikov and Bochvar.)
- Mostowski, A.** (1957) 'On a Generalization of Quantifiers', *Fundamenta Mathematicae* 44: 12-36. (Introduces 'generalized' quantifiers as an extension of first-order logic.)
- Nagel, E.** (1939) 'The Formation of Modern Conceptions of Formal Logic in the Development of Geometry', *Osiris* 7: 142-224; repr. in *Teleology Revisited and Other Essays in the Philosophy and History of Science*, New York: Columbia University Press, 1982, 195-259. (The influence of geometry on modern logic.)
- Padoa, A.** (1901) 'Essai d'une théorie algébrique des nombres entiers, précédé d'une introduction logique à une théorie déductive quelconque', in *Premier Congrès International de Philosophie*, vol. 3, *Logique et Histoire des Sciences*, Paris: Armand Colin, 309-65; partial trans. in van Heijenoort (1967), 118-23. (Method for showing a term to be undefinable.)
- Passmore, J.** (1968) *A Hundred Years of Philosophy*, Baltimore, MD: Penguin. (A general history emphasizing logic, especially good on Oxbridge logicians.)
- Peano, G.** (1888) *Calcolo geometrico*, Turin: Bocca. (Includes Peano's first contribution to logic, which was used in geometry.)
- Peano, G.** (1889) *Arithmetices principia, nova methodo exposita*, Turin: Bocca; partial trans. 'The Principles of Arithmetic', in J. van Heijenoort (1967), 83-97. (Includes the Peano postulates for the natural numbers.)

- Peano, G.** (1890) 'Démonstration de l'intégrabilité des équations différentielles ordinaires', *Mathematische Annalen* 37: 182-228.(First use of logical symbolism outside of arithmetic and geometry.)
- Peano, G.** (1891) 'Principii di logica matematica', *Revista di matematica* 1: 1-10; trans. 'Principles of Mathematical Logic', in *Selected Works of Giuseppe Peano*, Toronto, Ont.: University of Toronto Press, 1973, 153-61.(Peano's first full logical system.)
- Peano, G.** (1897) 'Logique mathématique', in *Formulaire de mathématiques*, Turin: Bocca, vol. 2, §1. (Peano's revised logical system.)
- Post, E.L.** (1921) 'Introduction to a General Theory of Elementary Propositions', *American Journal of Mathematics* 43: 163-85; repr. in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematics, 1879-1931*, Cambridge, MA: Harvard University Press, 1967, 264-83.(First published proof of the completeness of propositional logic; compare with Bernays (1918).)
- Russell, B.A.W.** (1903) *The Principles of Mathematics*, Cambridge: Cambridge University Press; 2nd edn, London: Allen & Unwin, 1937; repr. London: Routledge, 1992.(First publication of Russell's paradox and his logicism.)
- Russell, B.A.W.** (1906) 'The Theory of Implication', *American Journal of Mathematics* 28: 159-202.(His full development of propositional logic.)
- Russell, B.A.W.** (1922) 'Introduction', in L.J.J. Wittgenstein, *Tractatus Logico-Philosophicus*, trans. C.K. Ogden and F.P. Ramsey, London: Routledge; trans. D.F. Pears and B.F. McGuinness, London: Routledge, 1961. (Russell's introduction of a hierarchy of metalanguages.)
- Scanlan, M.** (1991) 'Who Were the American Postulate Theorists?', *Journal of Symbolic Logic* 56: 981-1,002. (Early American axiomatic theory.)
- Schröder, E.** (1877) *Der Operationskreis des Logikkalküls*, Leipzig: Teubner.(Schröder's first logical system.)
- Schröder, E.** (1890) *Vorlesungen über die Algebra der Logik (Exakte Logik) (Lectures on the Algebra of Logic (Exact Logic))*, vol. 1, Leipzig: Teubner; repr. New York: Chelsea, 1966.(Summary of the Boolean tradition in logic.)
- Schröder, E.** (1898) 'On Pasigraphy: Its Present State and the Pasigraphic Movement in Italy', *The Monist* 9: 44-62.(Schröder's logicism.)
- Tarski, A.** (1930) 'Über einige fundamentale Begriffe der Metamathematik', *Comptes rendus des séances de la Société des Sciences et des Lettres de Varsovie* 23: 22-29; trans. J.H. Woodger (1956), 'On Some Fundamental Concepts of Mathematics', in *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*, ed. J. Corcoran, Indianapolis, IN: Hackett Publishing Company, 2nd edn, 1983.(Tarski's first venture into metalogic.)
- Tarski, A.** (1952) 'Some Notions and Methods on the Borderline of Algebra and Metamathematics', in *Proceedings of the International Congress of Mathematicians, Cambridge, Massachusetts*, Providence, RI: American Mathematical Society, vol. 1, 705-20.(Beginnings of modern model theory.)
- Tarski, A.** (1956) *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*, trans. and ed. J.H. Woodger, Oxford: Clarendon Press; repr. and ed. J. Corcoran, Indianapolis, IN: Hackett Publishing Company, 2nd edn, 1983.(Translations of his papers on logic.)
- Veblen, O.** (1925) 'Remarks on the Foundations of Geometry', *Bulletin of the American Mathematical Society* 31: 121-41.(Veblen's plea for involving mathematicians with logic.)
- Whitehead, A.N. and Russell, B.A.W.** (1910) *Principia Mathematica*, vol. 1, Cambridge: Cambridge University Press; 2nd edn, 1925; repr. London: Routledge, 1994.(Their full development of the theory of types.)
- Wiener, N.** (1914) 'A Simplification of the Logic of Relations', *Proceedings of the Cambridge Philosophical Society* 17: 387-90.(Shows that ordered pairs, and hence relations, can be defined in the theory of types.)

Logic machines and diagrams

By 'logical diagrams' we generally mean any two-dimensional representations of logical relationships, such as of class inclusion or consequence. One usually also means representations using non-typographical symbols or geometric figures. Such diagrams were first used in the seventeenth and eighteenth centuries, but gained wide currency only in the nineteenth; the best known are the Euler and Venn diagrams. It is an open question whether logical diagrams are useful only as elementary pedagogical devices, or have implications for advanced logical research.

The conception of an organism, the mind, or of the universe as 'machine' was not really attractive and useful until machines were widespread, complex and able to perform interesting tasks. This occurred first in the late Renaissance, and initiated ways of thinking that dominated the seventeenth and eighteenth centuries. It was also becoming evident that machines could be used to perform some complex, repetitive or difficult tasks more reliably or faster than human beings.

The very idea of machines that can perform 'symbolic' tasks, such as mathematical, logical or, eventually, linguistic ones required first a symbolism. For this reason, the idea of computers for mathematical or logical tasks, and systems of mathematical and logical notation, are strongly intertwined: one must have efficient ways of feeding information into a machine, and interpreting the results.

1 Logical notation

Among the higher-level purposes for any symbolism in logic are: to abbreviate or simplify the writing or the display; to highlight strictly logical relationships without non-logical distractions (roughly Leibniz's *lingua philosophica* or his *characteristica universalis*); or to facilitate insights into the ways these representations may be manipulated, such as through inference (roughly Leibniz's *calculus ratiocinator*, 'calculus of reason').

It is customary to divide notational devices between those that represent substantive (often called 'categorematic') terms and those that represent logical relations or properties (often called 'syncategorematic' terms). Symbols for categorematic terms can stand for determinate individual or abstract entities, such as 'c' for Julius Caesar, 'a' for the class of apples, 'A' for the property of being an apple, or 'P' for the sentence 'All apples are red'. Symbols may also stand for arbitrary, non-specific examples of such terms, such as in the example, 'Consider a class of objects, c, or a proposition, P'. These are usually called variables, schemata or place-holders.

Perhaps the earliest use of abbreviations for categorematic terms was by Aristotle, who occasionally used single letters to stand for non-specific categorical terms in his *Prior Analytics*. While many later Aristotelian logicians occasionally used letters to represent categorical terms, only in the late medieval and early modern period do we find any use of special symbols, other than Latin or other 'natural language' expressions, for syncategorematic (logical) notions. One of the first of these was Luis Vives' use in the first half of the sixteenth century of a V-shaped symbol to represent term inclusion:

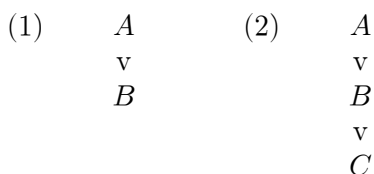


Diagram (1) represents that the term A is wholly included in B. ('All A's are B's.') Diagram (2) indicates that A is included in B, B is included in C, and, implicitly, A is (thus) included in C.

2 Philosophies of logical notation

The final goal of any logical notation is presumably to best allow one to carry out and communicate logical work. With Vives' and later notations, we begin to see four intermediate goals concerning how logical notation can achieve this final goal:

(1) The use of linear linguistic symbols, primarily deriving from natural language or other common usage (such as

mathematics) and used mainly as abbreviations. Polish notation is an extreme example of this phenomenon, using, for example, ‘ Kpq ’ for $p \ \& \ q$. (We might also include here mnemonic aids: for example, ‘K’ for Konjunktion - alas, only for speakers of some languages.)

(2) The use of special symbols chosen mainly for their distinctness from natural language symbols.

(3) The use of special symbols or notations that are ‘ideographic’ (Styazhkin) or ‘iconic’ (Peirce): properties of the symbols themselves are intended to mirror attributes of the represented logical relationship.

(4) Two- (or higher) dimensional notational systems using elements of 1, 2, or 3; that is, diagrams.

Observe that (1) and (4) may clash: achieving (1) may aid writers, printers and proof-readers, while achieving (4) may aid readers of the notational system. Similarly, (1) obviously clashes with (2) and perhaps with (3), if the natural language symbols are not fortuitously iconic (‘and’ is a non-palindrome although conjunction is symmetric).

Leibniz was fascinated by ideographic notations (owing to a passing interest in Chinese pictograms), but it was C.S. Peirce who first seems to have consciously chosen and systematically discussed symbolisms according to their iconic characteristics. One of his points is that to represent a non-symmetric logical relationship, one should use a visually non-symmetric symbol. That is, since $A \supset B$ does not always have the same truth-conditions as $B \supset A$, then one should use an asymmetric symbol, such as $\supset$, $\supsetneq$ or his own $\lhd$. Such considerations have been at work implicitly in the choice of virtually all logical symbols for conjunction, disjunction and so on (except those chosen in line with (1), such as ‘K’ and ‘Čamp;’).

Various symbols that came into widespread use in algebra and calculus in the sixteenth, seventeenth and eighteenth centuries, such as ‘=’ and ‘+’, involved similar (implicit) considerations, and hark back to suggestions by Viète, Descartes and Leibniz. The ‘algebra of logic’, such as we see in works by Leibniz, Lambert and, especially, Boole, borrows symbols already in use from algebraic notation (and thus is a species of goal (1)), but capitalizes upon their original iconic character in mathematics - goal (3). Both Peirce and Schröder later gave up the use of symbols for logical constants that were simply borrowed from algebra, using considerations from both (2) and (3), while preserving the algebraic (substitutional) character of many operations. (See [Logic in the 19th century](#).)

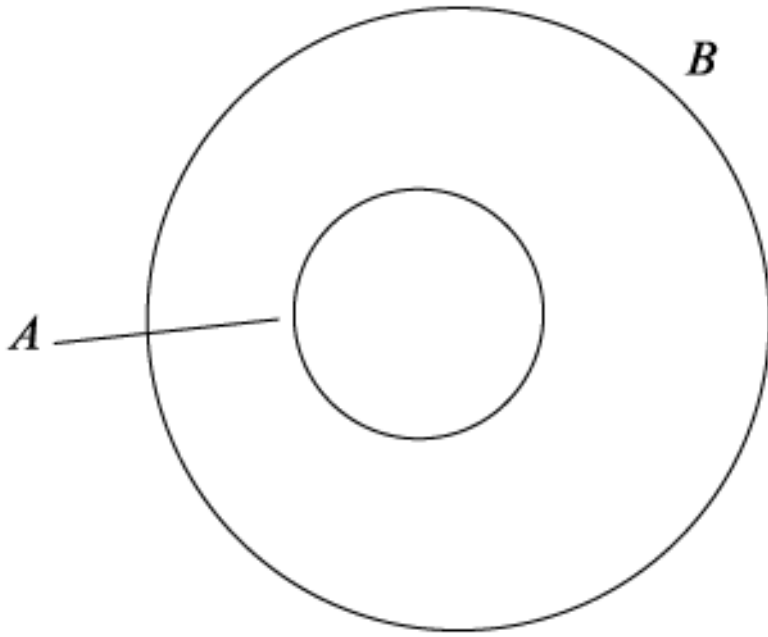
While the use of parentheses as indicating priority or scope of operations is of long-standing mathematical use, many of the other symbols now in use derive from the work of Giuseppe Peano. (These were sometimes inspired by earlier works in the algebra of logic, and were later injected into the mainstream of logic by Russell, Whitehead and others.) These symbols were primarily chosen by Peano according to consideration (2) - distinctiveness - and according to their suggestive, or etymological character: ‘ $\in$ ’ for the Greek ‘esti’; ‘ $\supset$ ’, a backward ‘C’, for the converse of ‘containment’, and so on. He seems not to have considered the ‘iconic’ arguments of Leibniz and Peirce, and forcefully rejected the more original two-dimensional proposals of Frege.

3 Euler and Venn diagrams

The two most widely used styles of logical diagrams are usually referred to as ‘Euler’ and as ‘Venn’ diagrams. Both are primarily used to represent inclusion or subsumption of classes or terms in Aristotelian logic. There are two main distinctions between these diagrams. Euler’s method involves placing circles of various sizes in various patterns; Venn’s uses circles of uniform size in fixed positions, which are then modified by other markings. Euler’s method does not have any unambiguous way of indicating the existence of individuals satisfying a certain description, other than implicitly by the spacing of the circles. Venn has such techniques.

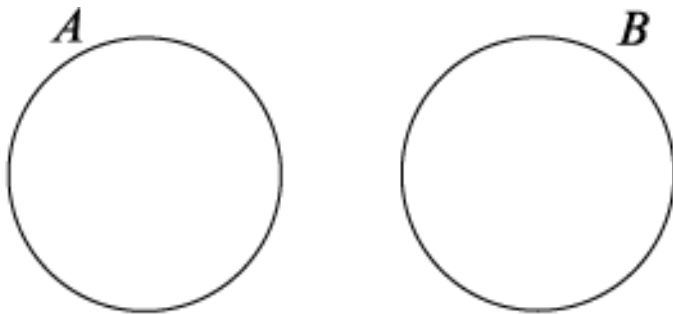
Leonhard Euler developed his method in his *Lettres à une Princesse d’Allemagne sur divers sujets de physique et de philosophie* (*Letters to a German Princess on Various Subjects of Physics and Philosophy*) in 1761, published in 1772 (see [Logic in the 17th and 18th centuries](#) §5). The method was later adapted by Gergonne, Maimon and others, and was frequently given an intensional as well as the original extensional interpretation. The following diagrams exhibit the method:

Euler Diagrams



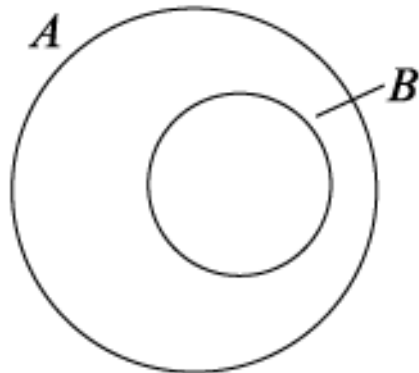
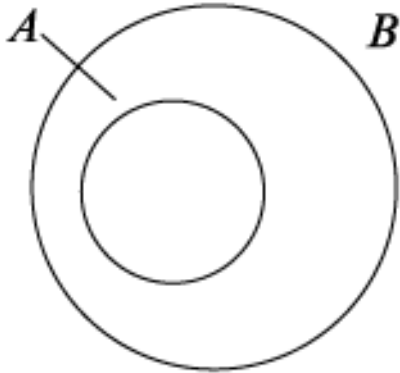
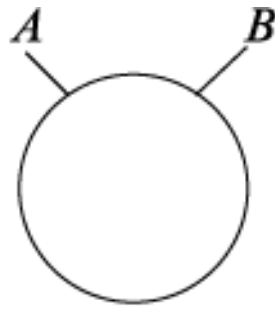
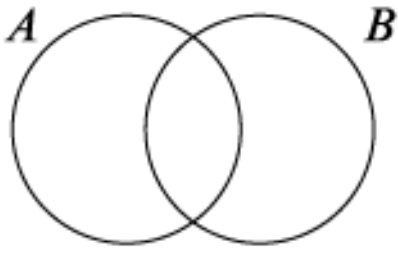
1. *A*-statement

Intensional reading: all the properties in concept *A* are included in the concept of *B*; all *A*'s are *B*'s. Extensional reading: all of the *A*-type things fall in the class of *B*-type things; all *A*'s are *B*'s.



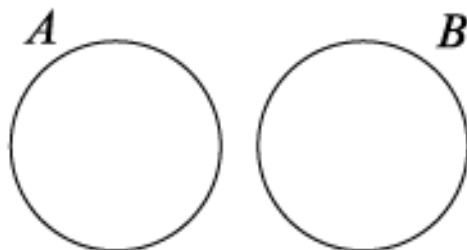
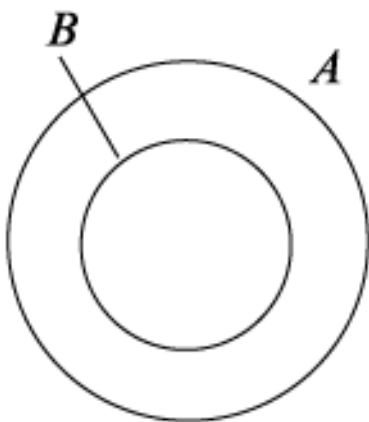
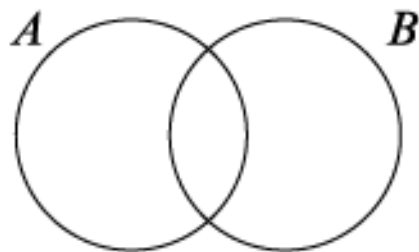
2. *E*-statement

Intensional reading: concepts *A* and *B* share no properties. Extensional reading: no individual is both *A* and *B*.



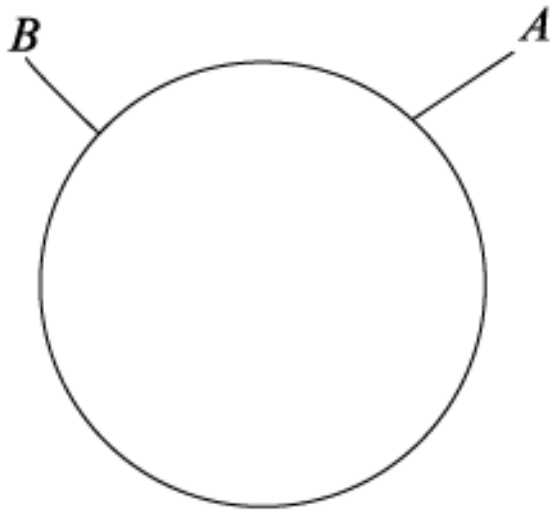
3. *I*-statement

Intensional reading: concepts *A* and *B* share some properties. Extensional reading: there are individuals that are both *A* and *B*. (Alternatives classified by Gergonne.)



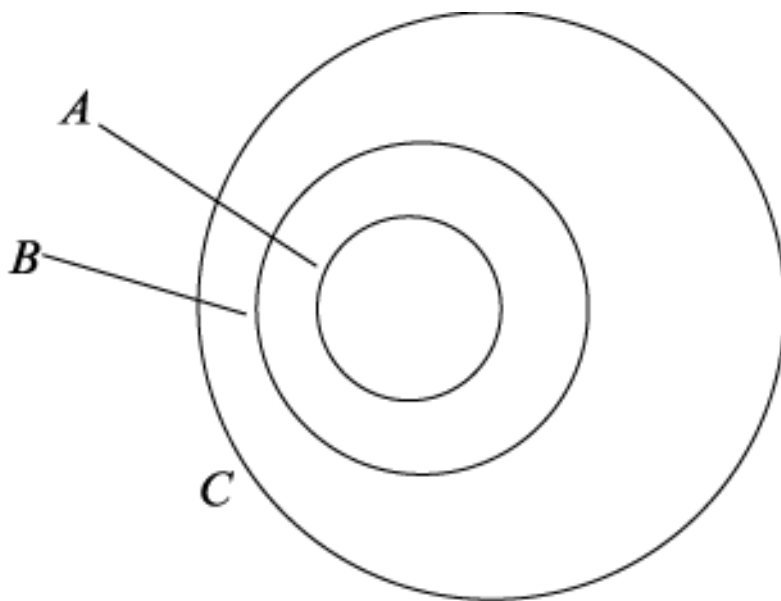
4. *O*-statement

Intensional reading: there are properties in concept *A* that are not in concept *B*. Extensional reading: there are individuals that are *A* but not *B*. (Alternatives classified by Gergonne.)



5. Identity

Intensional reading: *A* and *B* share all their properties. Extensional reading: the individuals that are *A* are the same as those that are *B*.



6. 'Barbara' syllogism

All *A*'s are *B*'s, all *B*'s are *C*'s, so all *A*'s are *C*'s, extensionally interpreted.

Observe that the intensional and extensional reading of the *A*-statement are 'reversed': to say that all apples are fruit is to say that all components of the concept of fruit (come from a plant, contain a seed, are edible, and so on) are included in the more numerous components of the concept of 'apple'. Observe also that there is an overlap in the alternative diagrams of the *O*- and *A*-statements (made explicit by Gergonne). Such an observation led to William Hamilton's (and George Bentham's) notorious doctrine of the 'quantification of the predicate', although similar ideas were common in the eighteenth and nineteenth centuries: we should distinguish between 'All *A*'s are some *B*'s' (1, 3, and 4) and 'All *A*'s are all *B*'s' (5).

The method of using movable, usually intensionally interpreted, circles was first suggested in published works by J.C. Sturm in 1661, and in unpublished writings of Leibniz. Contemporaneous with Euler, J.H. Lambert (1764) developed a method of using overlapping lines, also extensionally interpreted:

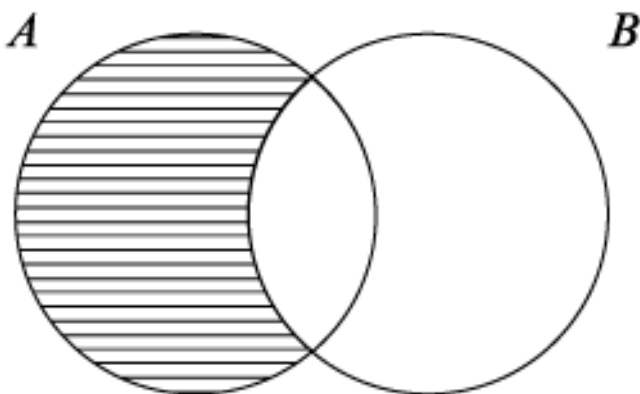
1. $A: \_ B: \_$
2. $A: \_ B: \_$
3. $A: \_ B: \_ \text{ or } \_ \text{ or } \_ \text{ or } \_$
4. $A: \_ B: \_ \text{ or } \_ \text{ or } \_$
5. $A: \_ B: \_$
6. $A: \_ B: \_ C: \_$

This method had also been anticipated by Leibniz in unpublished notes. In fact, the uses of circles or lines iconically to represent logical relationships reappears so often, and apparently independently, in the history of logic since the seventeenth century that one can count the idea - like that of an 'algebra' of logic - as more or less obvious to anyone who thought much about logic.

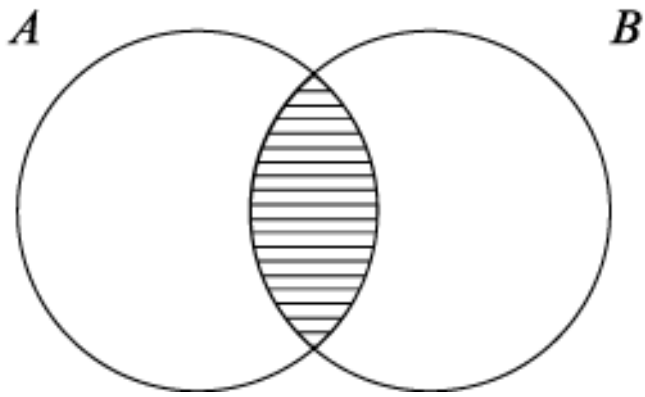
Venn diagrams are usually credited to John Venn in an article of 1880, and then more extensively developed in his *Symbolic Logic* (1881). This method was anticipated by similarly extensional diagrams of the German logician Moritz Drobisch in his 1838 *Neue Darstellung der Logik (New Conception of Logic)* (see *Logic in the 19th century §2*). After a decline in usage in the early twentieth century, Venn diagrams have reappeared in many introductory logic texts of the second half of the twentieth century.

Venn Diagrams

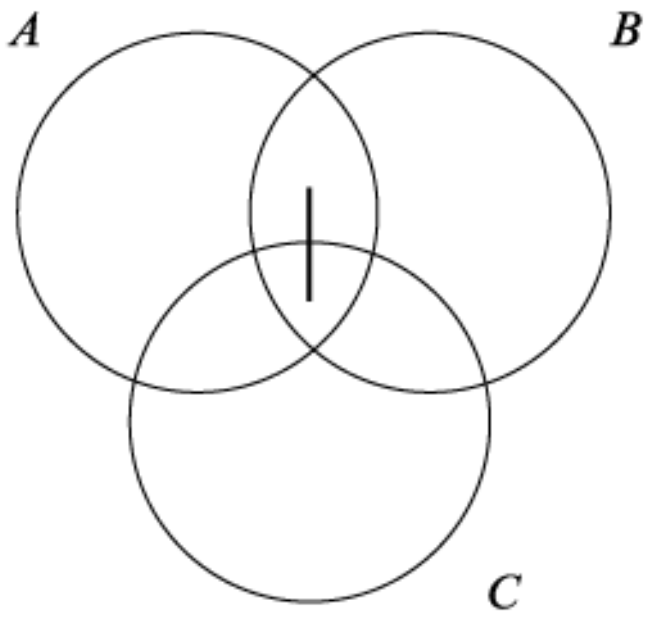
Statements are interpreted extensionally: points in an area indicate individuals having certain properties. What properties individuals have (or what classes they belong to) is indicated by inclusion in a circle: if they lack a property, they fall outside of that circle. The non-existence of individuals with a certain description is indicated by shading the corresponding area. The existence of an individual with certain properties (or lacking certain properties) is shown by an 'x' if it is fully determinate what properties it has, or by a line across two or more areas if it is not determinate.



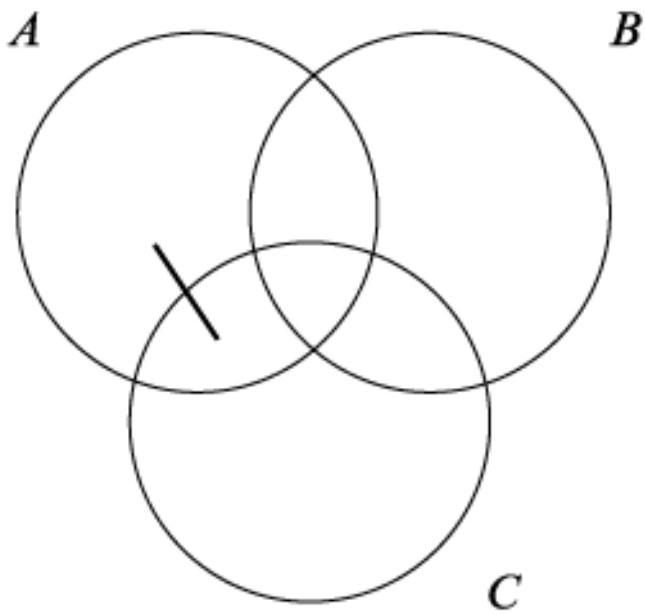
All A's are B's.



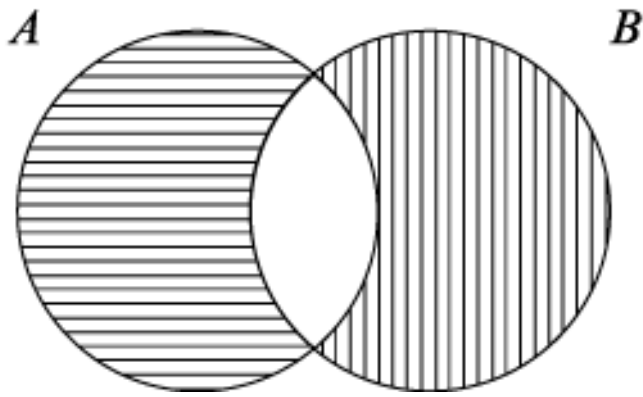
No *A*'s are *B*'s.



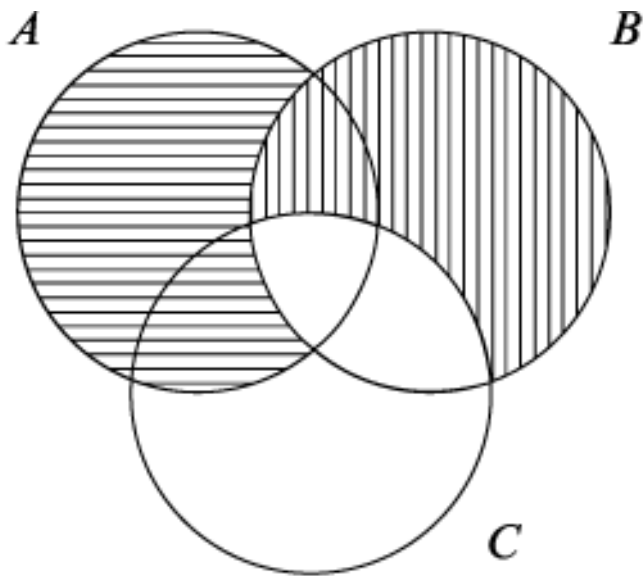
Some *A*'s are *B*'s: maybe they are *C*'s, maybe not.



Some *A*'s are not *B*'s.



All A 's are B 's, and all B 's are A 's.



All A 's are B 's.

All B 's are C 's.

So, all A 's are C 's.

The details of this interpretation are suggested by Drobisch and Venn, but carefully stated by Schröder in his *Vorlesungen* (1890-1905). Other diagrammatic methods in the nineteenth century included a system of triangles used by William Hamilton, and the use of matrices or tables by Allan Marquand and Lewis Carroll. Still other methods in the twentieth century include Veitch diagrams (see Veitch 1952) and Karnaugh maps. All of these methods, with the exceptions of the methods of Peirce and Frege, are versatile enough only for Aristotelian (monadic predicate) or propositional logic, and not for full first-order predicate logic.

4 Frege's *Begriffsschrift*

Gottlob Frege's notation, the famous '*Begriffsschrift*' ('conceptual notation'; see Frege, G.), is partly diagrammatic. He draws the material conditional in a two-dimensional way, writing $P \supset Q$ as follows.



The only specifically two-dimensional feature seems to be intended to highlight a feature of *modus ponens*, the main rule of inference using the material conditional - namely, the condition (antecedent) ‘drops away’. The use of the negation line and the descending vertical line (negation stroke) in the material conditional was clearly intended to suggest the interpretation of the material conditional $P \supset Q$ as $\neg P \vee Q$. Frege stubbornly stuck to his own notational system throughout his life, in spite of the difficulty it presented for typographers, criticism from Peano (and Schröder), its non-iconic and ahistorical character, and the fact that even admirers such as Russell and, later, Wittgenstein and Hilbert, never adopted it. Nevertheless, something of what motivated Frege, namely a concern for displaying features of *inferences* in a two-dimensional way, rather than the logical content or ‘sense’ of individual sentences, leads one to helpful formats for displaying mathematical or logical demonstrations (‘proofs’), such as the later ‘natural deduction’ methods of Gerhard Gentzen. (See [Natural deduction, tableau and sequent systems §1](#)).

5 Peirce’s existential graphs and conceptual graphs

There is little suggestion in any of these proposals that diagrams have any point much beyond a heuristic value, being ‘suggestive’, and helping to learn logic at the introductory level. No one seems to have proposed that diagrams have a serious, advanced use for professional logicians, or represent more accurately an underlying metaphysical reality. A notable exception are Charles S. Peirce’s ‘existential graphs’ (see [Peirce, C.S. §8](#)). Peirce, a historian of logic and a careful reader of his contemporaries, knew about the diagrammatic methods of Euler, Lambert and Venn. In the 1890s he developed three sophisticated systems of logical diagrams that he called the alpha, beta and gamma existential graphs. (His choice of ‘graph’ for these diagrams is somewhat unfortunate, since they are not graphs in the technical sense first used by J.J. Sylvester in 1878, now in wide use, and which Peirce would have known.) Both graph theory and Peirce’s graphs were inspired, interestingly, by chemical diagrams. Furthermore, other chemical metaphors occur in both the works of Frege (‘saturated’) and Peirce. Alpha graphs are similar to Venn diagrams, and can be used for either categorical logic or for what we would call propositional logic. Beta graphs can be used to represent any quantified relational sentence or set of sentences, and are thus a fully general method for diagramming sentences in first-order predicate logic - the first such method in the history of logic. Gamma graphs are used for the representation of modal statements. All three graph systems are ‘iconic’ in a sense of which Peirce was explicitly aware; he was also interested in enhancing their calculational usefulness for producing new inferences (‘forward inference’), or for identifying the correctness of given conclusions from given premises (‘backward inference’).

Furthermore, the graphs are not just employed for their usefulness in some narrow sense, but grew out of views he had developed much earlier: his ‘philosophy of notation’ of 1885. This was the first statement of what became a large, developed theory of signs (his semiotics) and a philosophy of mind - of thoughts themselves, at least in all a priori disciplines, being ‘diagrammatic’. (Peirce’s views are reminiscent of the better-known ‘picture theory’ of Wittgenstein’s *Tractatus*.) Peirce believed that our visual notations should match or facilitate, as much as possible, the best possible mental diagrams, and this means that the notational system should exhibit useful logical relationships. In earlier, algebraic-style writings, this was developed within a standard linear sequence of symbols; in later writings, he seems to have believed that two-dimensional representations were required for the properly ‘iconic’ display of some logical relationships, especially once relations are introduced.

Peirce’s ideas about existential graphs were not much noticed until the 1960s, and have not had a noticeable impact, except as a curiosity, until the 1980s. In computer science and artificial intelligence, Peirce’s ideas have been put forward as ‘conceptual graphs’ (by John Sowa 1991) and have found a wider audience. This is in part because there had existed in artificial intelligence a prior interest in using ‘semantic networks’ for representations of the meaning of natural language sentences. In cognitive science and learning theory, there is also a resurgence of interest in optimizing the display of information to native human cognitive processes, and thus of presenting complex relational information through visual images.

6 Early logic machines and mechanical machines for Aristotelian logic

Interestingly, the very first speculations about or design of machines that performed conceptual tasks - one hesitates to call them exactly ‘logic’ - were inspired not by arithmetic calculators, but by a strange mixture of medieval magic and theology. This was the fantastic *Ars Magna* (‘Great Art’ or ‘Technique’) of the Spanish

(Majorcan) mystic and tinkerer **Ramon Llull** (§2). Llull was interested in the way the various concepts (God, evil, ...) could be combined. To this end, he designed various notations, as well as 'machines' that used labelled, movable, concentric wheels to show these combinations. Although his views were respected by a variety of later thinkers such as Bruno and Swift, his greatest impact was on the development of the combinatoric theory by Pascal and Leibniz in the late seventeenth century.

Not surprisingly, the thought occurred to Pascal, Leibniz and others of using the combinatoric wheels of Llull, but labelling them with numerals, and gearing them so that they could do various arithmetic calculations. There are thus two conceptual steps that were necessary to produce the idea of logic machines. First, one must be able to think of machines as being able to perform calculations - to compute. (This in turn requires a notation and machines of some complexity and accuracy, that is, many finely machined parts.) Second, one must be able to think of reasoning and inference as a type of 'mechanical' calculation: this was the contribution of Llull, refined by Pascal and Leibniz, and explicitly stated by Leibniz as the *calculus ratiocinator*.

Both Pascal and Leibniz designed and had constructed for them geared arithmetic calculators. Leibniz also proposed, perhaps for the first time, the idea of a logic machine, and left some suggestions about how this might be done - but never developed precise designs nor constructed an actual machine.

The first reasonably successful logic machine was probably the eighteenth-century 'Demonstrator' of Charles Stanhope. It used a notation system that anticipated William Hamilton's variant of Aristotelian logic, and could also perform elementary arithmetic and probabilistic calculations. The Demonstrator was constructed from a block into which two moving slides were inserted, one of grey wood and the other of red glass; the positions of these slides served as the input premises and the extent of their overlap, visible in a small square in the block, represented the strongest possible conclusion validly derivable from these premises. That is, the Demonstrator was a 'forward inference' machine that actually generated a conclusion, rather than determining the validity or invalidity of a given conclusion with given premises. The use of the Demonstrator for deductive logic was limited to categorical logic with two premises.

The high-water mark of mechanical logic machines was reached by William Stanley Jevons' machine, usually called the 'Logical Piano', from the appearance of its keyboard. While restricted to categorical logic, this 1870 machine represented a vast improvement over Stanhope's Demonstrator, both logically and mechanically. Consisting of a keyboard for inputting premises and, above it, connected by pulleys and levers, a display of terms, the Jevons machine could accommodate up to four terms, and thus arguments with more than two premises ('sorites'), with designs for machines with up to ten terms. It could also handle 'or' terms in either subject or predicate terms (understood inclusively), such as 'All fruit are apples or oranges'. It used a Boolean equational system, and thus could also be considered the first algebra machine, since it essentially solved systems of equations. Like the Demonstrator, one did not input a conclusion to determine its validity given a set of input premises, but instead the machine output all valid conclusions with the input premises in the logical system.

7 Electrical machines and the twentieth century

A slight improvement in Jevons' design was made by Allan Marquand, who had been a logic student of Peirce and later became a professor of art history at Princeton. He called his a 'New Logical Machine' (1885). It was not logically more sophisticated than Jevons' machine, but had a better display, consisting of a grid of rotary indicators, looking something like a water or gas meter. Marquand also sketched designs for the first electrical logic machine, using relays and switches. (The very first electrical circuit diagram to be used for logic or interpreted logically was proposed in a letter from Peirce to Marquand.) Nothing really came of these proposals, and they, like Babbage's ideas for a general-purpose symbolic manipulator, were historical curiosities that had no discernible influence on electrical and electronic computers of the mid- and late twentieth century.

While there were a number of authors who developed mechanical logic machines in the late nineteenth and early twentieth centuries, mostly in imitation of Jevons' device, the collapse of the Aristotelian approach to logic made most of these techniques obsolete. Indeed, the logic of Frege, Russell and Whitehead, which permits quantifiers and relations, and which we know as predicate logic, makes the very idea of logic machines a daunting task. Expressions no longer consist of certain restricted combinations of terms. Furthermore, as Church's theorem was later to suggest, it may be that certain logical tasks in this framework are mechanically impossible. Similarly, the

idea of allowing a machine to reason whatever it can from given premises - forward inference, imitative of human reasoning - becomes extremely problematic. Designs for electrical logic machines appeared again shortly after the Second World War, first with work by two students of W.V. Quine, William Burkhart and Theodore Kalin in 1947, a number of machines in the late 1940s and early 1950s, and culminating in the 'Logic Theory Machine' of Allen Newell and Herbert Simon in 1956. (During the Second World War, Claude Shannon had developed his influential logical interpretation of electrical circuitry, thus rediscovering the observations of Peirce and Marquand.) Significantly, all these machines were restricted to propositional logic, and thus also fell far short of the expressivity and complexity of quantificational logic. Logic machines and computer programs in the twentieth century have been largely focused on 'backward' inference: confirming that a given conclusion does, or does not, validly follow from a given set of premises.

The history of logic machines grew distinct, after the early days of Leibniz and Pascal, from the history of machines to perform arithmetic calculations. There were enormous economic pressures to develop numerical calculators (for accounting and also for navigational tables), and few incentives to develop machines that reasoned using logic. Another line of development that was initially distinct from that of logic machines was the development of general purpose, programmable computers. Most early calculating and logic machines of all sorts were one-purpose and 'hard-wired'. Charles Babbage is the originator of the conception of a general-purpose 'computer', but the approach was rediscovered in the twentieth century, and intensively developed in mechanical, electrical and electronic implementations during and after the Second World War. In this approach, we build a 'computer' so that it can be 'programmed' to perform any task of symbolic manipulation that is capable of algorithmic description. Although the idea of such a general-purpose symbol-manipulator is an old one, going back to Babbage in the early nineteenth century, no one seems to have thought of programming such machines to perform logical tasks until Turing, nor succeeded until the Logic Theory Machine. Today, virtually all computational projects in logic are performed on general-purpose computers, and thus in an important sense the idea of dedicated 'logic machines' is obsolete. It is the software (programming), not the hardware, that is designed to perform logical tasks. Starting in the 1960s and accelerating rapidly in the 1970s and 1980s, a large amount of computer software (including 'mechanical theorem provers') now exists that performs a variety of logical tasks in various logical systems, including first- and second-order predicate logic, 'deviant' logics and modal logics.

See also: [Computability and information](#); [Computability theory](#); [Logic and mathematical terms, glossary of](#)
 RANDALL R. DIPERT

References and further reading

- Cajori, F.** (1952) *A History of Mathematical Notations*, La Salle, IL: Open Court; repr. New York: Dover, 1993. (A classic book - and the only one - on a monumentally important topic in intellectual history; nevertheless dated and much in need of a treatment using recent historical, logical and cognitive-science research.)
- Eames, C. and Eames, R.** (1990) *A Computer Perspective: Background to the Computer Age*, ed. G. Fleck, Cambridge, MA: Harvard University Press, new edn. (A pictorial and well-researched history of computers that includes pictures, diagrams and facsimiles of Stanhope's, Jevons' and Marquand's logic machines, Peirce's circuit diagrams, and some post-Second World War work.)
- Gardner, M.** (1982) *Logic Machines and Diagrams*, Chicago, IL: University of Chicago Press, 2nd edn. (A classic discussion of the topic, and the only monograph, but somewhat elementary and superficial.)
- Goldstine, H.** (1972) *The Computer from Pascal to von Neumann*, Princeton, NJ: Princeton University Press, 1993. (The history of arithmetic and general-purpose computers, including some 'logic' machines.)
- Kneale, W. and Kneale, M.** (1962) *The Development of Logic*, Oxford: Clarendon Press, 1984. (Standard but now dated work in the history of logic, with some brief discussions of notations and diagrams, and mention of some logic machines.)
- Lambert, J.H.** (1764) *Neues Organon oder Gedanken über die Erforschung und Bezeichnung des Wahren und dessen Unterscheidung vom Irrthum und Schein (New Organon or Thoughts on the Investigation and Designation of the True and its Distinction from Error and Appearance)*, Leipzig, vol. 1, part 1, 109-33; repr. in *Philosophische Schriften*, vols 1-2, ed. H.W. Arndt, Hildesheim: Olms, 1965. (Introduces the diagrams sometimes known as 'Lambert lines'; also referred to in later works by Venn and John Neville Keynes in the nineteenth century.)
- Peano, G.** (1973) *Selected Works of Giuseppe Peano*, trans. and ed. H. Kennedy, Toronto, Ont.: University of Toronto Press. (A small collection of Peano's huge output, this includes some of his important writings on

logical and mathematical notations, and his visionary proposals for a new universal language, ‘Latin without inflections’.)

- Peirce, C.S.** (1931-58) *Collected Papers*, ed. C. Hartshorne, P. Weiss and A.W. Burks, Cambridge, MA: Harvard University Press, 8 vols; and *Writings of Charles S. Peirce: A Chronological Edition*, ed. M. Fisch *et al.*, Bloomington, IN: Indiana University Press, 1982-. (Important logical papers of 1867-86 appear in volume 3 of the *Collected Papers*; other volumes include some writings on the existential graphs and iconicity.)
- Roberts, D.D.** (1973) *The Existential Graphs of Charles S. Peirce*, The Hague: Mouton. (The only book-length work in English on Peirce’s existential graphs; more accessible than Peirce’s own works.)
- Schröder, E.** (1890-1905) *Vorlesungen über die Algebra der Logik*, Leipzig: Teubner, 3 vols; repr. New York: Chelsea, 1966. (The monumental *locus classicus* of the algebra of logic, including relations; extensive discussions of diagrams, various notations, and philosophical aspects of them.)
- Shin, S.J.** (1990) *The Logical Status of Diagrams*, Cambridge: Cambridge University Press. (Situation theory is used to provide a precise analysis of reasoning with Venn diagrams.)
- Sowa, J.** (ed.) (1991) *Principles of Semantic Networks*, San Mateo, CA: Morgan Kaufmann. (Discussions of various network and graph formalisms in cognitive science, including the Peirce-inspired conceptual graphs.)
- Styazhkin, N.I.** (1964) *History of Mathematical Logic from Leibniz to Peano*, Cambridge, MA, and London: MIT Press, 1969. (Often more useful discussions of diagrams than Kneale and Kneale (1962), including Lambert lines.)
- Veitch, E.W.** (1952) ‘A Chart Method for Simplifying Truth Functions’, *Proceedings of the Association for Computing Machinery* 1952: 127-33. (An exposition of what have come to be known as ‘Veitch diagrams’.)
- Venn, J.** (1881) *Symbolic Logic*, London: Macmillan, 2nd edn, 1894; repr. New York: Chelsea, 1979. (By the time it appeared, this work was unoriginal, but introduced in textbook fashion the now-popular Venn diagrams.)

Logic, medieval

Medieval logic is crucial to the understanding of medieval philosophy, for every educated person was trained in logic, as well as in grammar, and these disciplines provided techniques of analysis and a technical vocabulary that permeate philosophical, scientific and theological writing. At the practical level, logic provided the training necessary for participation in the disputations that were a central feature of medieval instruction, and whose structure - with arguments for and against a thesis, followed by a resolution - is reflected in many written works. At the theoretical level, logic, like other subjects, involved the study of written texts through lectures and written commentaries. The core of the logic curriculum from the twelfth century onwards was provided by the logical works of Aristotle. These provided the material for the study of types of predication, the analysis of simple propositions and their relations of inference and equivalence, the analysis of modal propositions, categorical and modal syllogisms, fallacies, dialectical Topics, and scientific reasoning as captured in the demonstrative syllogism. Comprehensive as this list might seem, medieval logicians realized that other logical subjects needed to be investigated, and, again from the twelfth century onward, new techniques and new genres of writing appeared. The main new technique involved the use of '*sophismata*', or puzzling cases intended to draw attention to weaknesses and difficulties in logical definitions and rules. The new genres of writing especially included works on '*supposition theory*', which concerned the types of reference that the subjects and predicates of propositions have in different contexts, and works on '*syncategoremata*', which concerned the effect on sense and reference produced by the presence and placing of such logical terms as 'all', 'some', 'not', 'if...then', 'except', and so on. Other important topics for investigation include '*insolubles*', or semantic paradoxes, and '*consequences*', or valid inference forms. These new developments were seen as providing a supplement to Aristotelian logic, rather than an alternative. The only context in which people occasionally suggested that Aristotelian logic was inapplicable was that of Trinitarian theology, and the only logician who deliberately set out to reform logic as a whole was Ramon Llull.

The study of medieval logic involves two kinds of difficulty. In the first place, few texts are available in translation, and indeed, many are not even available in printed form. In the second place, there is a problem of interpretation. For a very long time, the specifically medieval contributions to logic were ignored or despised, and when people began to take them more seriously, there was a strong tendency to look at them through the spectacles of modern formal logic. More recently, scholars have come to realize that medieval interests cannot be mapped precisely onto modern interests, and that any attempt, for example, to make a sharp distinction between propositional and quantificational logic is misleading. The first task of the modern reader is to try to understand what the medieval logician was really concerned with.

1 Background

The story of medieval logic is the story of logic in western Europe after the split of the Roman Empire into the Latin-speaking, Catholic west and the Greek-speaking, Orthodox east (see [Byzantine philosophy §3](#)). Because this split was followed by barbarian invasions and constant upheaval, very little happened between 550 and 1000, and learning was mainly kept alive in Benedictine monasteries. However, it is important to consider the Latin texts that were transmitted, for they formed part of the basis for later developments. One very important early source is the *Peri Hermeneias* by Apuleius of Madaura which dates from the second century AD. It discusses the proposition and the categorical syllogism, and it was influential until the late twelfth century. It was the main source for Martianus Capella (early fifth century), author of *De nuptiis Philologiae et Mercurii* (*The Marriage of Philology and Mercury*), Cassiodorus (sixth century), author of *Institutiones*, and Isidore of Seville (seventh century), author of *Etymologies*, all of whose encyclopedic works include sections on logic (see [Encyclopedists, medieval §§4, 6, 7](#)). Other early works include Marius Victorinus on definitions and a work attributed to Augustine, the *Categoriae decem* (*The Ten Categories*), which is from the fourth-century circle of Themistius and was the most intensely studied logical work in the ninth and tenth centuries.

Most important of all are the works of Boethius. He seems to have been responsible for translating all of Aristotle's *Organon* (that is, his six works on logic) into Latin, and all but the *Posterior Analytics* survive. He also translated Porphyry's *Isagōgē* (*Introduction*), and wrote commentaries on some of Aristotle's logic, on Porphyry and on the *Topics* of Cicero. In addition he composed monographs of his own on Division, on Topics, on

categorical syllogisms and on hypothetical syllogisms, that is, on conditional propositions and arguments built up from them (see [Boethius, A.M.S. §§2-3](#)). At first, little attention was paid to this body of writing, but signs of revival first appear at Charlemagne's court, for which Alcuin (d.804) wrote a little work on dialectic (see [Carolingian renaissance §2](#)). The first outstanding figure is [Gerbert of Aurillac](#), who taught in the cathedral school at Reims from about 972 before becoming a bishop and finally Pope. He taught the *Logica vetus* (*Old Logic*) (that is, Porphyry's *Isagōgē*, Aristotle's *Categories* and *On Interpretation*), Cicero's *Topics* and a good deal of Boethius.

As western Europe became more settled, so educational institutions began to flourish. Better monastic schools (for example, Bec, where Anselm taught) and cathedral schools (see [Chartres, School of](#)) developed, and in the twelfth century, especially in Paris, schools grew up around individual masters such as Abelard and Adam of Balsham [Parvipontanus]. The monographs and commentaries of Boethius had been largely recovered by the beginning of the eleventh century and, with the *Logica vetus*, became the focus of attention. The *Liber sex principiorum* (*Book of Six Principles*) (that is, the last six categories, about which Aristotle had had less to say), attributed to Gilbert of Poitiers, became part of the *Logica vetus*. At the same time the philosophical study of grammar became important, with a focus on the *Institutiones grammaticae* (*Institutions of Grammar*) of Priscian (see [Language, medieval theories of §1](#)). Anselm's important discussion of the reference of the word 'grammaticus' ('expert in grammar' or 'literate') shows an attempt to harmonize Aristotle's and Priscian's doctrines (see [Anselm §5](#)). By far the most outstanding figure, and the first really original logician, was Peter [Abelard \(§§2, 4\)](#). Unfortunately, his own turbulent life and the important events of the later twelfth century combined to diminish any influence his logical work may have had.

2 After 1150: a general survey

Logic after 1150 is shaped by four interrelated developments: the recovery of the rest of Aristotle's logical works, along with other texts; the foundation of universities and the *studia* of the new religious orders; the development of 'sophismata', a new technique for handling logical problems; and the development of new areas of logic.

New translations. Aristotle's *Topics* and *Sophistical Refutations* were known by the 1130s and the entire *Logica nova*, including the *Prior* and *Posterior Analytics*, was known by 1159 when John of Salisbury referred to them in his *Metalogicon*. All circulated in translations by Boethius, with the exception of the *Posterior Analytics*, which was translated by James of Venice in the twelfth century. Both the *Topics* (see [Alexander Neckham](#)) and the *Sophistical Refutations* were used in the twelfth century, but most especially the latter, and the *Prior Analytics* gradually displaced Boethius' works on the categorical syllogism. The *Posterior Analytics* was absorbed more slowly. The first known complete commentary is that by [Grosseteste \(§1\)](#), dating from the early 1220s.

In the second half of the twelfth century Arabic logic began to appear. The first part of Avicenna's *Logic* (see [Ibn Sina §1](#)), dealing with the *Isagōgē*, was translated by Gundissalinus, as was Farabi's *Catalogue of Sciences*, whose second chapter dealt with logic. The *Logic* of Ghazali was also translated. In the 1230s the logic commentaries of Averroes on the *Categories*, *On Interpretation* and the *Prior* and *Posterior Analytics* were translated, though they were less successful than the Arab works translated earlier. Some Greek commentators were also translated. Themistius' *Paraphrase of the Posterior Analytics* and some extracts from Philoponus on the *Posterior Analytics* became known in the late twelfth century. Much later William of Moerbeke translated Simplicius on the *Categories* (1266) and Ammonius on *On Interpretation* (1268). The frequently cited but now lost commentary by 'Alexander' on *Sophistical Refutations* was probably the translation of a work by Michael of Ephesus (c.1130). Other Greek works on logic were not translated until the Renaissance. The important thing about all these texts is not just their content but the fact that they provide a full curriculum for an organized institution. (See [Aristotelianism, medieval §3](#) for the twelfth-century reception of translations.)

New institutions. By the beginning of the thirteenth century the Universities of Oxford, Paris and Bologna were taking shape as organized institutions. Gradually other universities were founded, along with the *studia* of the new teaching orders, especially the Dominicans (at whose Cologne *studium* Albert the Great taught) and the Franciscans (at whose London *studium* Ockham taught), and all these institutions offered a training in the liberal arts (see [Encyclopedists, medieval §1](#)). Two things should be noted. First, a degree from the arts faculty was required for all advanced (graduate) study in theology, medicine and law (though especially in Italy, medicine and law, unlike theology, could be pursued at the undergraduate level). Second, logic formed a very large part of the

arts curriculum, especially during the first two years. One reason for its importance was the large place given to disputation at all levels of university teaching: logic clearly offers training in the techniques of debate, and inculcates quickness in analysing and responding to arguments. However, university teaching also involved the close study of texts, especially the texts of Aristotle, and many commentaries, as well as original textbooks, were written by medieval logicians. (See [Natural philosophy, medieval §5](#) for teaching practices.)

Sophismata: a new technique. Medieval logic is characterized by a new technique, the analysis and solution of sophismata. The word ‘sophisma’ covers two phenomena. First, there is the sophisma sentence itself, which is a logical puzzle intended to introduce or illustrate a difficulty, a concept, or a general problem. The origin of these sentences seems to lie in the twelfth-century use of ‘*instantiae*’, or counter-examples. Second, there is the technique of the sophismatic disputation, which is used to show that the very same reasoning which supported a plausible thesis could also be used to establish something implausible. By the end of the twelfth century the sophisma was established in different genres of logical and grammatical writing. These genres included treatises on syncategoremata, that is, those words which perform a function in a sentence by negating, conjoining, disjoining and so on (see §8), and treatises on sophismata themselves. The relationship between these treatises was very close. Typically, a treatise on syncategoremata would start with a syncategorematic term and then appeal to sophismata to illustrate the difficulties its use might cause. A treatise on sophismata would start with a sophisma and, using disputational techniques, appeal to facts about a syncategorematic term or to logical distinctions to solve it, or to show that different truth-values were possible according to different senses of the sophisma sentence. Various types of sophismata and treatises on sophismata survive. It seems that in Paris the sophisma was an element of teaching, and had to be debated in the schools, that is, within a fairly formal setting. Very elaborate sophismata survive from the latter part of the thirteenth century which were given solutions by the teaching masters, for example, [Boethius of Dacia](#), and which were often used as a vehicle for the straightforward discussion of interesting logical problems. On the face of it, Oxford sophismata seem to have been quite different. They were the subject of live debate, but this debate was at a strictly undergraduate level, and was part of the practical training in logical disputation. They were not primarily a vehicle for discussing doctrine, and they seem to disappear as a genre when the curriculum changes. Fourteenth-century Oxford sophismata had two special features. First, they exhibit a new emphasis on the ‘*casus*’, or initial hypothesis about the context of the sophisma-sentence, which might itself be identified as the source of the problems in determining the truth-value of that sentence. This notion of *casus* played a key role in treatises on insolubles (see §10) and obligations (see §11), which are both closely linked to sophismata. Second, many of them are about natural philosophy, and introduce mathematical considerations and calculatory techniques (see [Oxford Calculators §§2, 4](#)).

Among the surviving fourteenth-century treatises on sophismata are important works by Richard [Kilvington](#), William [Heytesbury](#), John Buridan and Albert of Saxony. In the fifteenth century, [Paul of Venice](#) wrote on sophismata, as well as using them extensively in other works. Both the technique and the genre of writing disappeared early in the sixteenth century.

3 After 1150: new logical writings

While the writings of Aristotle were central to the logic curriculum, there were matters that he did not discuss, which left room for a considerable number of new developments. The most prominent is the so-called ‘*logica moderna*’ or ‘*modernorum*’ (a label coined by twentieth-century authors). Also known as ‘terminist logic’, this includes supposition theory and its ramifications (see §7). Treatises on supposition theory deal with the reference of subject and predicate terms in propositions, and they have as a corollary the treatises on syncategoremata, which deal with all the other terms in propositions. The material dealt with in treatises on syncategoremata was also dealt with in treatises on sophismata (see §2 above), but in the fourteenth century much of it also appeared in treatises on the ‘proofs of terms’ (see §8). Three more important developments are found in treatises on consequences (see §9), insolubles (see §10) and obligations (see §11).

While supposition theory constitutes the most important new development, it was not always privileged. Paris in the second half of the thirteenth century, and Bologna and Erfurt at the beginning of the fourteenth century, were centres of ‘modist logic’, and tended not to use supposition theory. Modist logicians (for example, [Peter of Auvergne](#)) were more interested in the sense than the reference of terms, and they made much use of the relationships between the *modi essendi* (modes of being) of things, the *modi intelligendi* (modes of understanding)

and the *modi significandi* (modes of signifying) of words. They were also characterized by their lengthy discussions of the rational nature of logic and its relation to second intentions (see §4). Another rival to supposition theory as a tool of analysis was provided by ‘proofs of terms’, which were particularly important in the later fourteenth century.

Another new form of writing was the comprehensive textbook. At least six survive from the thirteenth century, including those by William of Sherwood, Peter of Spain and Roger Bacon. In the fourteenth century we find those by William Ockham, John Buridan and Albert of Saxony. Some universities, especially Oxford and Cambridge, preferred to use loose collections of brief treatises on various topics, and a good example of such a collection is the *Logica parva* of Paul of Venice.

There are three things to note about these new writings. First, they were not officially part of the curriculum at most universities, but that did not prevent their being used. Second, because Latin was the common language of instruction in western Europe, a text written in one place could be used anywhere else. Peter of Spain’s text was generally the most popular, but Buridan was particularly important in Poland, and in late fourteenth- and fifteenth-century Italy short texts from England were heavily used. Third, there is no genuine division between the authors of Aristotle commentaries and the authors of the new texts. To take just one example, John Buridan wrote a comprehensive textbook, commentaries on parts of Aristotle’s *Organon*, and separate treatises on consequences and sophismata, the last containing a noteworthy treatment of insolubles.

4 The nature of logic

The purpose of logic had nothing to do with the setting up of formal systems or the metalogical analysis of formal structures. Instead, it had a straightforwardly cognitive orientation. Everyone accepted the view that logic is about discriminating the true from the false by means of argument. This is why logic was essential to the speculative sciences, since it provided the instruments for finding truth and for proceeding from the known to the unknown.

The first problem with logic was what to call it. The name ‘*dialectica*’ was most prevalent until the thirteenth century, when ‘*logica*’ gained the upper hand. Both words have various associations. ‘*Dialectica*’ in the broad sense just is logic, but the word also has three narrower senses: dialectic as the art of debate; dialectic as the art of finding material for arguments; and dialectic as a kind of reasoning which falls between demonstrative and sophistical reasoning. The first sense was appealed to by Roger Bacon when he said that logic as an art has to do with disputation. The second sense is associated with the discussion of Topics, the headings under which the material for arguments can be sorted. Because the study of Topics also included maxims, or self-evidently true generalizations, which could provide the warrant for different types of argument, there is a close link between Topics and argumentation. Hence, the third and most usual sense of *dialectica* had to do with topical or dialectical syllogisms as a part of logic. Medieval logicians treated Aristotle’s distinction between dialectical and demonstrative syllogisms as an epistemological one concerning the status of their premises, so that dialectical syllogisms had the same formally valid structure as demonstrative syllogisms.

The problems associated with the word ‘*logica*’ have to do with whether logic is a linguistic or a rational pursuit. Isidore of Seville noted that ‘*logica*’ comes from the Greek word ‘*logos*’ which can mean ‘*sermo*’ (word) or ‘*ratio*’ (reason). As a result logic could be called either a *scientia sermocinalis* (linguistic science) or a *scientia rationalis* (rational science). There were considerations supporting both titles. On the one hand, the Stoics had divided philosophy into natural, moral and rational, and the last was equated with logic which could then, as Boethius pointed out, be seen as both an instrument and a part of philosophy (see [Stoicism](#)). On the other hand, logic was one of the liberal arts (see [Encyclopedists, medieval §1](#)) and belonged to the *trivium*, along with rhetoric and grammar, which made it seem a linguistic science. This emphasis was intensified by the discovery of Arab logicians who included Aristotle’s *Rhetoric* and *Poetics* in the *Organon*, or logical works, a classification accepted by Albert the Great and Aquinas, among others. Some logicians, such as William of Sherwood, preferred to call logic just a linguistic science, but many others in the thirteenth century, including Robert Kilwardby and St Bonaventure, called it both linguistic and rational.

In the late thirteenth and fourteenth centuries the notion of logic as a rational science only became predominant. This move was partly associated with the rediscovery of the *Posterior Analytics*, the new emphasis on demonstrative science, and the increased importance of the classification of science into practical and speculative,

along with the associated division of speculative science into natural philosophy, mathematics and divine science. Unlike the Stoic classification, this one offered no obvious place for logic. Some later logicians, including Ockham, Buridan and Albert of Saxony, classified logic as a practical science, but most preferred to think of it as speculative (though of course with practical applications). People then called logic a supporting science (*scientia adminiculativa*), subordinated to the three important sciences. The notion that logic was a tool directing reason in the acquisition of knowledge was one reason for calling it rational.

However, the most important reason for calling logic rational had to do with views about its subject matter, and here the influence of Avicenna is crucial, for it was Avicenna who had said that logic was about second intentions. Second intentions, often identified with beings of reason (*entia rationis*), are those higher level concepts we use to classify our concepts of things in the world, and they include such notions as genus, species, subject, predicate and syllogism. There were obvious ontological difficulties here, for one could argue that second intentions pick out special common objects including both universals and logical structures, and one could also argue that second intentions are just mental constructs reached through reflection on individual things and on actual pieces of discourse or writing. Nominalists and realists obviously disagreed on ontology, but this did not prevent such nominalists as Ockham from agreeing that logic deals with second intentions, and that the syllogism the logician considers is neither a thing in the world, nor a piece of writing or speaking. Some people preferred to say that logic was about things in the world as they fall under second intentions, and other people preferred to pick out some special second intention such as argumentation or the syllogism as the subject of logic, but there was still a strong consensus that the objects of logic are rational objects.

These remarks have to be balanced by the obvious fact that people reason in ordinary language, and that ordinary locutions are often vague, ambiguous or misleading. Medieval logicians spent an enormous amount of time on the analysis of ordinary Latin, not because they thought that logic was theoretically concerned with ordinary language, but because this was the only way to avoid fallacious reasoning. Indeed, the avoidance of fallacy is at the heart of all the new types of logical writing.

5 Syllogistic

Despite the new forms of logical writing, syllogistic remained basic, and the syllogism was generally regarded as the central form of argumentation, to which all other forms could be reduced. This view is a lot less restrictive than one might imagine, because of the general way in which ‘syllogism’ was defined. Following Aristotle, logicians wrote: ‘A syllogism is an expression in which, when certain things have been asserted, something else must occur by means of the things which were asserted’ (Peter of Spain, in [Kretzmann and Stump 1988: 217](#)). This definition encompassed modal syllogisms, syllogisms with singular terms (called ‘expository syllogisms’), tensed syllogisms with past and future verbs, syllogisms with ‘oblique’ (or inflected) terms such as genitives (for example, ‘Every man’s donkeys are running, Brownie (the donkey) is a man’s, therefore Brownie is running’), and syllogisms containing such exponible terms (see §8) as ‘only’ and ‘except’.

The standard core of syllogistic included immediate inference as captured by the square of opposition, the rules of equipollence (for example, ‘Not every man is running’ is equivalent to ‘Some man is not running’), the laws of conversion and contraposition, and the categorical syllogism itself, together with the laws of reduction which allow one to prove non-first-figure syllogisms on the basis of the first-figure syllogisms, treated as axioms, and some other rules including conversion. (For details, see [Logic, ancient](#).) The main medieval contribution here was the provision of mnemonic devices, particularly the verse ‘Barbara, Celarent,...’, which accounts for nineteen valid categorical syllogisms, together with detailed rules of reduction (see Peter of Spain, in [Kretzmann and Stump 1988: 224-5](#)).

One topic of discussion was whether a fourth figure should be recognized. This question relates to the ways in which the major and minor terms can be defined. From the sixteenth century onward, it became more usual to adopt the definition of Philoponus and to say that the major term is the predicate of the conclusion, while the minor term is the subject of the conclusion. This definition gives four figures which represent four possible arrangements of the major, minor and middle terms in two differentiated premises. Alternatively, if one adopts the standard medieval definition, whereby the major term is the term that appears with the middle term in the first premise, one can still speak of four figures with differentiated premises but it turns out that each figure also has indirect modes, in which the major term is the subject of the conclusion. The normal medieval account encompassed just the direct

and indirect modes of the first figure, and the direct modes of the second and third figures. People argued that the fourth figure is just the first indirect figure with transposed premises, and they did not count this as involving a genuinely different disposition of the middle term. John Buridan seems to have been the first to realize (in rather brief remarks) that on the medieval account, one will still have four distinct figures, and that the direct modes of the first figure will count as the indirect modes of the fourth figure in just the same way as the direct modes of the fourth figure count as the indirect modes of the first.

Another subject of discussion concerned the relation of perfect and imperfect syllogisms. ‘Perfect’ syllogisms were first-figure syllogisms whose necessity was obvious, and so needed no proof in terms of simpler argument forms, though they can be said to be founded on the regulative principles *dici de omni* (which ‘occurs when nothing is to be subsumed under the subject of which the predicate is not said’) and *dici de nullo* (which ‘occurs when nothing is to be subsumed under the subject from which the predicate is not removed’): Peter of Spain, in [Kretzmann and Stump 1988: 217](#)). Links between *dici de omni* and the Topic of quantitative whole were occasionally mentioned in the thirteenth century, but the general view was that perfect syllogisms do not need to be justified by appeal to Topics. Other syllogisms, however, are imperfect, and need to be justified by the rules of reduction, which involve the use of the laws of conversion. This immediately raises the problem of circularity, for if all arguments are reducible to the syllogism, it seems to follow that the laws of conversion are themselves reducible. There was considerable uncertainty on this point, and Burley, in his early treatise on consequences, says both that everything is reducible to the syllogism and that the laws of conversion are not. One way out of the impasse was to see the syllogism itself as just one among many forms of consequence (see §9), but although this is the approach taken by the later Burley and by Buridan, followed by Albert of Saxony, it is less prevalent than has often been suggested. Nor is there any reason to suppose that Burley or Buridan regarded themselves as showing that quantificational logic is based on propositional logic when they suggested that general inferences with unanalysed parts have some kind of priority.

The laws of contraposition raised an interesting semantic issue, which was discussed by Buridan and by various other logicians up to the early sixteenth century. It was commonly held that affirmative sentences with a non-referring subject were false, and the corresponding negative sentences true. But if one accepts term-negation while holding that sentences with term-negation (and no other) are affirmative, contraposition of the sort ‘All men are beings, therefore all non-beings are non-men’ will be invalid, given that there are no non-beings. Buridan and other late medieval logicians responded by claiming that the laws of contraposition could hold only when an extra existential premise was supplied. It is important to note a corollary of their view. In classical quantificational logic, the standard interpretation of sentences with non-referring terms is that universal sentences (formulated with ‘if...then’) are true, and particular ones (formulated with ‘and’) are false. As a result, quantificational formulations of the simple inferences ‘All *A*s are *B*, therefore some *A*s are *B*’ and ‘No *A* is *B*, therefore some *A*s are not *B*’ are rejected as invalid, as are the quantificational formulations of syllogisms with universal premises and particular conclusions. When medieval logicians had to decide on a uniform way of treating sentences with non-referring terms they decided not to draw a line between universal and particular sentences, but between affirmative and negative ones. As a result, the simple inferences captured in the square of opposition, the laws of equipollence, the laws of conversion and the standard categorical syllogisms remain valid. It is only the laws of contraposition that need reformulation.

6 Compounded and divided senses; modal logic

The discussion of modal logic is in part parallel to the discussion of categorical syllogistic. In general introductions to logic the discussion of modal propositions was followed by the discussion of simple modal inferences and equivalences as captured in a square of modal opposition, modal conversion and, finally, modal syllogisms. Buridan is noteworthy for the large number of modal equivalences he considered, and Ockham for his thorough exploration of different kinds of modal syllogism. Modal logic also figures in treatises on consequences, on proofs of terms, and on compounded and divided senses. The central modal notions employed were ‘necessary’, ‘impossible’, ‘possible’ and ‘contingent’, of which the last two need some explication. In the earlier period such logicians as Peter of Spain followed Boethius in treating ‘possible’ as equivalent to ‘contingent’, but later logicians, such as Buridan, made a careful distinction between ‘possible’ as compatible with ‘necessary’ but not with ‘impossible’, and ‘contingent’ as incompatible with both ‘necessary’ and ‘impossible’.

A full understanding of modal logic requires a consideration both of the basic tool of analysis, the distinction between compounded and divided senses, and also of the general theory of modality.

The distinction between compounded and divided senses has its origin in Aristotle's discussion of the fallacy of composition and division, to which Abelard was one of the first to pay careful attention. The basic point concerns two ways of reading the sentence 'A seated man can walk'. Interpreted according to its compounded sense, this proposition is *de dicto* (about a dictum or 'that' clause - see [De re/de dicto §1](#)) and means 'That-a-seated-man-walk (that is, while seated) is possible'. Interpreted according to its divided sense, the proposition is *de re* (about a *res* or thing), and means 'A seated man has the power or ability to walk'. The proposition is false in the first sense, but true in the second. It became standard when considering modal inferences in general and modal syllogisms in particular to distinguish between the compounded and divided senses of premises and conclusion, and to work out the logical results of these different readings.

In the fourteenth century the distinction between the compounded and divided senses was very widely used, as can be seen from the work of William Heytesbury, the first to devote a complete treatise to this topic. He listed nine types of logical problems which could be solved by paying attention to the distinction. These problems were signalled by the presence of modal words but also, for example, by the presence of future tense verbs, verbs producing confused supposition (see §7), terms such as 'infinite' which could be taken both categorematically and syncategorematically (see [Oxford calculators §§2, 4](#)), and words about acts of will and intellect. This last case ties in with epistemic logic, as a branch of applied modal logic. For instance, Heytesbury cites the inference 'Your father you believe to be a donkey; therefore, you believe your father to be a donkey' ([Kretzmann and Stump 1988: 426-7](#)), and remarks that it is invalid because the antecedent is taken in the divided sense ('Your father happens to be the thing [at a distance] that you take to be a donkey': true) and the consequent is to be taken in the compounded sense ('You believe that your father is a donkey': false).

The general theory of modality underwent interesting changes during the Middle Ages. In the earlier period, there were two possible approaches. On one approach, the interpretation of modalities is straightforwardly ontological and possibility is seen as a potency, an aspect of beings and events to be discussed in the context of change and motion. A second interpretation of modalities is the so-called statistical interpretation. On this account, to say that an object or state of affairs is necessary is to say that it occurs at all times, to say that it is possible is to say that it occurs at some but not all times, and to say that it is impossible is to say that it never occurs. For both these views, to talk about modalities is just to talk about what exists or occurs at various times, and modality is not some feature of the world over and above what there is. Moreover, they are tied to the principle of the necessity of the present, that is, that if *P* is true at time *t*, then it is not possible for not-*P* to be true at time *t*. An alternative model of possibility was at first confined to purely theological contexts, but in the twelfth century such authors as Gilbert of Poitiers and Abelard began to take a new approach, and in the fourteenth century Duns Scotus worked out the full philosophical implications of the alternative model, the so-called synchronic interpretation. On this model, possibility involves reference, not just to the present state of affairs, or to the potentialities of objects, but to states of affairs or worlds which are here and now alternative to the present world. There is no necessity of the present, and the basic notion is that of conceptual consistency. This new view leads to more complete and satisfactory formulations of the relations between modal sentences, especially when they are taken in the divided sense.

7 Supposition theory

The most notable new theory that took shape in the twelfth century was supposition theory and its ramifications, particularly ampliation and restriction. In the early period, *suppositio* was taken to be a property of a substantive term, while *copulatio* was a property of an adjectival term, and *appellatio* a property of a term that referred to actual existents. Later, the term *copulatio* was dropped, and *appellatio* acquired a totally new meaning (see below). Roughly speaking, supposition theory concerned what a categorical term (the subject or predicate of a proposition) can be taken to stand for in a particular context, and it was used to diagnose fallacies and test inferences.

The three main types of supposition were material, simple and personal. A term was said to have 'material' supposition when it stood for itself or its equiforms, as in 'Man is a noun'. Material supposition took the place of twentieth-century quotation devices, and gave rise to the same sorts of problems. A term was said to have 'simple' supposition when it stood for a universal, as in 'Man is a species', but this type was controversial, and not only because the question of the nature of universals is itself highly controversial. Peter of Spain suggested that

nouns used as predicates, such as ‘animal’ in ‘Every man is an animal’, had simple supposition, and William Arnaud in the late thirteenth century suggested that singular terms, such as ‘Socrates’ in ‘Socrates is an individual’, had simple supposition. These views were not generally accepted. Finally, a term has ‘personal’ supposition when it is taken for its normal referents, as when ‘man’ is taken for Socrates, Plato and so on.

Some logicians distinguished accidental personal supposition from natural supposition which allowed a term to have pre-propositional reference to all its referents, past, present and future, while others insisted that supposition must be purely propositional and contextual. This debate in turn affects the doctrines of ‘ampliation’, whereby the reference of a term can be extended, and ‘restriction’ (the opposite). Parisian logicians, such as Jean le Page (writing c.1235), tended to accept natural supposition, and to say (like Buridan in the fourteenth century) that terms had natural supposition in scientific propositions - that is, universal necessary truths - so that no ampliation was necessary. As a corollary, in non-scientific propositions the supposition of terms was restricted in various ways. For English logicians in the thirteenth century, all supposition was contextual, and the notion of ampliation had to be used when the subject of a proposition was to extend beyond present existent things. Both groups agreed that modal terms were particularly apt to produce ampliation. For instance, in ‘A man can run’, ‘man’ is amplified to supposit for present and future men (Lambert of Lagny, mistakenly called ‘of Auxerre’, in [Kretzmann and Stump 1988: 117](#)).

The three types of personal supposition most often appealed to are determinate, purely confused (*confuse tantum*), and confused and distributive. These types were normally illustrated by means of the descent to singulars. For instance, to say that the subject of a particular affirmative proposition, ‘Some *A* is *B*’, has ‘determinate’ supposition is to say that one can infer the disjunction of singular propositions, ‘This *A* is *B*, or that *A* is *B*, or the other *A* is *B*,...’. To say that the predicate of a universal affirmative proposition, ‘Every *A* is *B*’, has ‘purely confused’ supposition is to say that one can infer a proposition with a disjoint predicate, ‘Every *A* is this *B* or that *B* or the other *B*,...’. To say that the subject of a universal affirmative proposition has ‘confused and distributive’ supposition is to say that one can infer a conjunction of propositions, ‘This *A* is *B*, and that *A* is *B*, and the other *A* is *B*,...’.

Some people distinguished between mobile and immobile cases. For instance, no descent is possible from ‘Only every *A* is *B*’, and so *A* has ‘immobile’ supposition. A fourth type of supposition is ‘collective’ supposition, as in ‘Every man is hauling a boat’, given that they are doing it together. Here any descent will involve a conjoint subject, as in ‘This man and that man and the other man...are all hauling a boat’.

The theory of personal supposition was used to solve a variety of problems. Some involved tense (for example, how to convert the proposition ‘No old man will be a boy’); some involved multiple quantification (for example, how to distinguish ‘Every donkey of some man is running’ from ‘Some donkey of every man is running’, to take an early sixteenth-century example); and some involved such verbs as ‘promise’, as in ‘I promise you a horse’. How such sentences should be analysed, and what kinds of inferences they could figure in, elicited a variety of complex solutions. The positions adopted included these: Heytesbury took it that ‘horse’ had purely confused supposition; Paul of Venice added that the supposition was immobile; and Buridan appealed to the new doctrine of appellation, according to which some verbs appellate a form, that is, bring into play the notion of something coming under a certain description. Ockham preferred to replace the sentence by a more complex sentence, ‘You will have one horse by means of my gift’. This solution is closer to the type found in treatises on the proofs of terms, which focused, not on the reference of the term ‘horse’, but on the analysis of the word ‘promise’ in terms of giving a right to an object.

8 Syncategoremata; proofs of terms

Supposition theory concerned the reference of the subject and predicate terms in propositions; treatises on syncategoremata dealt with all the other terms, such as ‘all’, ‘some’, ‘not’, which appear in a proposition and exercise some logical function. The two areas of investigation were very closely associated. Not only are there terms such as ‘infinite’ which can be interpreted both as categorematic and as syncategorematic, but some terms were dealt with in each type of treatise. For instance, in the first half of the thirteenth century the word ‘all’ or ‘every’ (*omnis*) was discussed in the tracts on distribution associated with supposition theory in the textbooks of Jean le Page, Peter of Spain and Nicholas of Paris, but in the treatises on syncategoremata by the Englishmen William of Sherwood and Robert Bacon. Peter of Spain and Nicholas of Paris both wrote treatises on

syncategoremata in which ‘*omnis*’ is not discussed. As was noted above (§2) these works were very closely associated with those on sophismata.

Treatises on syncategoremata were most prominent in the thirteenth century, and they did not altogether disappear in the fourteenth. For instance, the late-fourteenth-century English logician, Richard Lavenham, wrote one. However, two other forms of writing come to the fore. First, there are a lot of short treatises on particular syncategorematic terms, including ‘know’ (*scire*), ‘begins’ (*incipit*) and ‘ceases’ (*desinit*), and terms with the power of producing purely confused supposition. Second, and most important, are the treatises on the proofs of terms (or proofs of propositions), whose best-known example is the *Speculum puerorum* of Richard Billingham, an Oxford author of the mid-fourteenth century.

In this context, a proof seems to be a method of clarifying a sentence containing a particular sort of term, or of showing how one might justify that sentence. There were three groups of terms. ‘Resoluble’ terms are those whose presence calls for explanation or clarification through ostensive reference, as captured in an expository syllogism (that is, one with singular terms). Thus ‘A man runs’ is resolved into the expository syllogism ‘This runs; and this is a man; therefore a man runs’. ‘Exponible’ terms are those whose presence calls for exposition of the sentence in terms of a set of equivalent sentences. For instance, the sentence ‘Only a man is running’, which contains the exclusive term ‘only’, is expounded as ‘A man is running and nothing other than a man is running’. Other exponible terms are exceptives, such as ‘except’, reduplicatives, such as ‘inasmuch as’, ‘begins’ and ‘ceases’, ‘infinite’, and so on. In fact, they are the terms which had figured prominently in treatises on syncategoremata. Finally, there are ‘official’ (*officiales*) or ‘officiable’ (*officiabiles*) terms, so called because they performed a function (*officium*). These included any term that governed a whole sentence or that treated a whole sentence as modifiable, such as modal terms, and such terms as ‘know’, ‘believe’, ‘promise’, ‘desire’ and ‘owe’. Analysis of sentences containing such terms shows why they are referentially opaque when taken in the compounded sense.

Treatises on proofs of terms were very popular during the late fourteenth and fifteenth centuries, but they were joined by treatises which dealt with exponible terms alone. Pierre d’Ailly wrote one such treatise, and several were written in the early sixteenth century.

9 Consequences

The notion of ‘*consequentia*’ was already discussed in the context of conditional statements by Garlandus Compotista, who wrote in the eleventh century, and was heavily influenced by Boethius’ work on hypothetical syllogisms and on Topics. However, it was not until the beginning of the fourteenth century that consequences became the subject of separate studies, one of the very earliest being written by Walter Burley around 1300. The origin of these treatises is controversial. It has been argued that consequences grew out of the study of Topics, especially topical maxims (see §4), but there is no sign of a theory of consequences in thirteenth-century works on Topics, and very little in thirteenth-century commentaries on the *Prior Analytics* or the *Sophistical Refutations*, both places where discussion might be expected. There is some evidence of relevant discussion in treatises on syncategorematic terms, especially under ‘*si*’ (if...then), but it seems most likely that ordered treatment grew out of the English desire to produce lists of rules for use in disputations about sophismata. This ties in with the strong orientation of many fourteenth-century English treatises towards syncategorematic terms. A small group of authors, including Ockham and the later Burley, did produce treatises between about 1325 and 1340 which are more oriented towards Topics, but these constitute a minority treatment. The rare Continental treatises, such as that of John Buridan (c.1335), had more to do with syllogistic than with syncategoremata. English treatises, especially those by such later fourteenth-century authors as Ralph Strode and Richard Ferribrigge, were used and commented on in the fifteenth century in Germany and Italy.

One problem to do with consequences is the relationship between a *consequentia* and a conditional statement. Unlike modern logicians who make sharp distinctions between syntax (what the rules allow one to infer) and semantics (validity), and between language (statements) and metalanguage (statements about statements), medieval logicians seem to have been indifferent about whether to present consequences as rules of inference, as arguments which may be valid or invalid, or as conditional statements. Buridan, for instance, defined a consequence as a hypothetical (that is, compound) proposition, and said that a good consequence could be called true or valid. The standard remark that all true conditionals are necessary, and all false ones impossible, strengthened the tendency to equate consequences and conditionals, as did the use of ‘antecedent’ to refer to the conjunction of the members of

the set of premises and ‘consequent’ to refer to the conclusion of a consequence.

Another problem concerns the definition of validity. The claim that a consequence is valid if and only if it is impossible for the consequent to be false when the antecedent is true was questioned for two reasons. In the first place, the propositions involved were taken to be occurrent items, whether written, spoken or mental. They could fail to exist, in which case there is nothing to carry a truth-value. Alternatively, their meaning could be at odds with their actual expression, as in ‘Every proposition is affirmative, therefore no proposition is negative’. Such problems were discussed at length, for example, by Buridan, who solved them by substituting a definition in terms of signification, that is, that a consequence is valid if and only if it is impossible for it to be as signified by the antecedent without its being as signified by the consequent.

The presence of ‘if and only if’ raises the second problem. The truth definition (or Buridan’s substitute) may provide a necessary condition for validity, but is it sufficient? If it is sufficient, then we must accept the paradoxes of strict implication, that is, that anything follows from an impossible proposition, and a necessary proposition follows from anything. The debate about these paradoxes began in the twelfth century, and there was a series of attempts to provide a second condition which, with the first, would be sufficient for validity. Abelard had a containment principle by which the dictum of the antecedent should contain the dictum of the consequent (see [Abelard, P. §4](#)), and Robert Kilwardby in the next century, like Strode in mid-fourteenth-century Oxford, said that the consequent had to be understood in the antecedent. Some people in the thirteenth century focused on reality, and argued that a consequence must capture a causal relation, and that as a result the antecedent must be about a state of affairs that can at least be supposed to be possible. None of these people could accept the paradoxes as formally valid. On the other hand, the Parvipontani or Adamites (followers of Adam of Balsham) in the twelfth century and Buridan in the fourteenth were happy to accept the first clause as offering both necessary and sufficient conditions for validity, with all that that implied for the acceptance of the paradoxes.

One favourite division of consequences, found in such authors as William of Sherwood and Robert Kilwardby, was into natural and accidental (or non-natural) consequences. A natural consequence met the two-clause condition for validity, and an accidental one did not. The paradoxes could then be classified as accidental consequences. Later authors tended to make a division into formal and material consequences, but we have to be careful here, as there were two approaches to the notion of material consequence. For some, a material consequence involved the independence of antecedent and consequent, and just was the two paradoxes. For others, such as Buridan, a materially valid consequence is defined as one which does not hold for all terms arranged in the same way, and includes such enthymemes as ‘If all animals are running, then all humans are running’.

Another division of consequences is into those valid absolutely (*simpliciter*) and those valid as of now (*ut nunc*). This division, which appears in Ockham, has been much discussed, though it is in fact relatively rare, at least in fourteenth-century English treatises. Consequences valid *ut nunc* have been thought to correspond to the modern material conditional, but this is not so, as modal notions are definitely involved. Given that Socrates may not exist, and given that sentences with non-existent subjects are held to be false, ‘If all men are running, then Socrates is running’ is not generally valid. However, if Socrates does exist, then it is impossible for the antecedent to be true and the conclusion false, and we have a consequence which is valid as of now.

The question of the paradoxes also leads us to the question of the relationship between consequences and propositional logic, as discussion of the paradoxes provided one of the main occasions for sequences of rules to be used in a proof. The typical sequence was as follows: from the contradiction ‘*P* and not-*P*’ we can infer *P*, and from *P* we can infer ‘*P* or *Q*’. From ‘*P* and not-*P*’ we can also infer not-*P*, and from ‘*P* or *Q* and not-*P*’ we can infer *Q*. Hence, from ‘*P* and not-*P*’ we can infer *Q*. This sequence is found in the twelfth-century theologian Alexander Neckham, and it reappears through the centuries. But the explanations given differ widely. In his treatise on syncategoremata, Peter of Spain appealed to topical maxims, and claimed more generally that we need topical maxims as a foundation for consequences. On the other hand, in Pseudo-Scotus, writing c.1350, we get an appeal to what seem to be straightforward rules of propositional logic.

It has been claimed that later treatments of consequences showed that medieval logicians grasped that quantificational logic is founded on propositional logic, but this is anachronistic. Texts on consequences contain a mixture of rules (from the point of view of the modern logician, that is). Some are general (‘If *A* is the antecedent of *B*, and *B* is the antecedent of *C*, then *A* is the antecedent of *C*’). Some are propositional and truth-functional

(‘Since a conjunction is true if and only if its conjuncts are true, from “A and B” we may infer A’). *Modus ponens*, *modus tollens* and De Morgan’s rules are given. Other rules are syllogistic; yet others are modal. Medieval logicians had no interest in classifying rules according to different formal systems, since they did not operate with the notion of a formal system in the first place.

10 Insolubles

‘Insolubles’ (*insolubilia*), so-called, it was claimed, because they are difficult to solve rather than completely insoluble, are semantic paradoxes. The simplest version is the liar paradox, ‘What I am saying is false’, given the *casus* or initial situation that this is all that is said, but complex versions with hypothetical propositions (‘God exists, and some conjunctive proposition is false’) or sequences of mutually referring propositions (‘Suppose that Socrates says “Plato says something false”, and Plato says “Socrates says something true”’ - Albert of Saxony, in Kretzmann and Stump 1988: 357, 349) were also discussed. In the twentieth century such paradoxes have been used to cast doubt on the very foundations of semantic theory, and have led to elaborate distinctions between levels of language and metalanguage. Medieval logicians, however, showed no signs of such a crisis mentality, and while they did employ certain restrictions on self-reference, and make certain distinctions between language and metalanguage, these techniques were generally limited to the problem in hand.

A very early formulation of the liar paradox is found in Adam of Balsham, writing in 1132, and later in the twelfth century Alexander Neckham gives a formulation. Full discussion seems to have started at least by the second half of the twelfth century, and continued until the sixteenth century. In the early period insolubles seem to have been discussed in the framework of Aristotle’s *Sophistical Refutations*, and after that, within the general framework of a sophismatic disputation. From about 1320 to 1350 there was a period of intense and sophisticated activity, during which new solutions were given, and the framework of an obligational disputation was used (see §11). After 1350 some authors, such as John Wyclif and Pierre d’Ailly, still had original solutions to offer, but most authors, including Paul of Venice, merely elaborated previous ones.

The early solutions to the paradox include ‘restriction’, which forbids self-reference, either by saying that a part (‘false’) cannot supposit for a whole (‘I say what is false’), or by saying that ‘I say’ cannot refer to the present moment; and ‘*cassatio*’, according to which a paradox says nothing at all. The first major new solution was that of Thomas Bradwardine (c.1320-4). Bradwardine claims that propositions signify whatever follows from them and that insolubles signify that they are true. As a result, an insoluble signifies that it is true, but it is true if and only if it is both true and false. This cannot be, so it is false. John Buridan (§2) and Albert of Saxony both follow Bradwardine. Roger Swyneshed was author of another popular solution. He said that truth requires not just correspondence with reality but also absence of self-falsification, and Paul of Venice gives a version of this solution in his *Logica Magna*.

11 Obligations

Treatises on ‘obligations’ discussed the rules for obligational disputations, which were logical exercises for students. They got their title because of the rules obliging one to answer in a certain way. Obligational terminology is found in the late twelfth century, and the theory was already well-advanced in Paris by 1240-50. In 1302 Walter Burley wrote a treatise which summarized much previous material, and laid the foundations for later developments. Other important treatises were by Roger Swyneshed (1330-5), Ralph Strode and Marsilius of Inghen. The terminology of obligational disputation, including the use of a *casus* or description of the initial situation, permeated much fourteenth-century writing, including theology.

Various types of obligational disputation were discussed. One classification, which has thirteenth-century origins and was popularized by Marsilius of Inghen in the second half of the fourteenth century, related to the three legitimate ways of responding: ‘*concedo*’ or ‘I grant it’; ‘*nego*’ or ‘I deny it’; and ‘*dubito*’ or ‘I doubt it’. Walter Burley had six types, including ‘*impositio*’, in which words and sentences were given new meanings for the duration of the disputation.

The most important of the different types of obligational disputation was ‘*positio*’. In this type of disputation, the opponent began the game by positing a proposition which the respondent had to grant, provided that it was logically possible. This *positum* was normally false, and it was usually trivial in content. Next the opponent had to

propose a series of other propositions, which could be either true or false. In replying to these, the respondent had to follow the rules of the game. If a proposition followed from the set formed by the *positum*, the propositions already granted and the negations of propositions already denied, it had to be granted. If a proposition was inconsistent with the set so determined, it had to be denied. If a proposition was such that neither it nor its negation followed from the set so determined, it was called *impertinens* or irrelevant, and the respondent had to reply to it according to his current state of knowledge. The task of the opponent in this game was to lead the respondent to the point of accepting an inconsistent set, either by granting and denying the same proposition, or by granting a proposition whose negation would follow from the set of propositions already granted. The task of the respondent, of course, was to avoid falling into these traps. In order to be successful, the respondent needed an excellent grasp of the logical relations between propositions and also an excellent memory. This is why obligational disputations could serve as valuable exercises for young logicians.

Some changes are reflected in the treatises. In the early period, impossible *positio* was very important, and still occurs in Ockham, though very rarely afterwards. In this kind of disputation, one begins with an impossible proposition which can at least be conceived to be true, a condition which excluded the use of explicit contradictions. Another change related to developments in modal logic. In the thirteenth century, the respondent had to deny that the actual 'now' of disputation was the instant at which the false but possible *positum* is treated as true, and as a result the replies to irrelevant propositions were related to actual time, while the replies to relevant propositions were related to an imaginary time. Once the notion of synchronic alternatives to the actual world had been adopted, all replies could be related to the same time. In the later period, Roger Swyneshed altered the relationship between irrelevant propositions and the *positum*. In the usual rules, once an irrelevant proposition had been admitted into the sequence, it (or its negation) became relevant. For Swyneshed, relevance was strictly a matter of a relation between the *positum* and those propositions that follow from or are incompatible with it alone. This produces two separate sets of responses, one involving relevant propositions, and one involving irrelevant propositions. As a result, one can deny a conjunction both of whose parts have been granted, and grant a disjunction both of whose parts have been denied. This view generated a good deal of discussion in the fourteenth and fifteenth centuries, and was usually rejected.

The biggest problem presented by treatises on obligations has to do with their purpose. They were certainly closely connected with consequences, insolubles and sophismata. Equally certainly they were closely connected with undergraduate training. Among other possibilities, it has been suggested that they also presented theories of counter-factual conditionals, or that they represented thought-experiments in which retrospective construction of a possible world takes place.

See also: [Language, medieval theories of](#); [Logic, Renaissance](#); [Logical and mathematical terms, glossary of](#)
E.J. ASHWORTH

References and further reading

- Ashworth, E.J.** (1978) *The Tradition of Medieval Logic and Speculative Grammar from Anselm to the End of the Seventeenth Century: A Bibliography from 1836 Onwards*, Toronto, Ont.: Pontifical Institute of Mediaeval Studies. (Annotated bibliography with full indexes.)
- Ashworth, E.J. and Spade, P.V.** (1992) 'Logic in Late Medieval Oxford', in J.I. Catto and R. Evans (eds) *The History of the University of Oxford*, vol. 2, *Late Medieval Oxford*, Oxford: Clarendon Press, 35-64. (Detailed account of the period 1330-1500 when English logicians were influential throughout Europe.)
- Broadie, A.** (1993) *Introduction to Medieval Logic*, Oxford: Clarendon Press, 2nd edn. (A brief, readable introduction to some parts of medieval logic. The first edition should not be used.)
- Buzzetti, D., Ferriani, M. and Tabarroni, A.** (eds) (1992) *L'Insegnamento della Logica a Bologna nel XIV Secolo (Logic Teaching at Bologna in the Fourteenth Century)*, Bologna: L'Istituto per la Storia dell'Università. (A useful collection of Italian studies and Latin texts related to logic teaching at an important Italian university.)
- Green-Pedersen, N.J.** (1984) *The Tradition of the Topics in the Middle Ages: The Commentaries on Aristotle's and Boethius' 'Topics'*, Munich and Vienna: Philosophia. (The standard study of Topics; from Aristotle to the sixteenth century.)
- Jacobi, K.** (1988) 'Logic (ii): The Later Twelfth Century', in P. Dronke (ed.) *A History of Twelfth-Century Western Philosophy*, Cambridge: Cambridge University Press, 227-51. (Excellent introduction to the debates in

the twelfth-century schools of logic.)

- Jacobi, K.** (ed.) (1993) *Argumentationstheorie: scholastische Forschungen zu den logischen und semantischen Regeln korrekten Folgerns* (*Theory of Argumentation: Scholastic Researches into the Logical and Semantic Rules of Correct Inference*), Leiden, New York and Cologne: Brill. (A collection of 35 studies, 25 in English, on inference in various parts of medieval logic, with full introductions to the different sections in both German and English. An excellent source, with a full bibliography.)
- Knuuttila, S.** (1993) *Modalities in Medieval Philosophy*, London and New York: Routledge. (A difficult but rewarding study of modal logic and how it was employed.)
- Kretzmann, N., Kenny, A. and Pinborg, J.** (eds) (1982) *The Cambridge History of Later Medieval Philosophy*, Cambridge: Cambridge University Press. (Much of this volume is devoted to logic between 1200 and 1350; the best single source on this period available.)
- Kretzmann, N. and Stump, E.** (trans.) (1988) *The Cambridge Translations of Medieval Philosophical Texts*, vol. 1, *Logic and the Philosophy of Language*, Cambridge: Cambridge University Press. (An invaluable collection of translations of authors from Boethius to Albert of Saxony; designed to supplement the [Cambridge History](#) (Kretzmann, Kenny and Pinborg 1982).)
- Lewry, P.O.** (1984) 'Grammar, Logic and Rhetoric 1220-1320', in J.I. Catto (ed.) *The History of the University of Oxford*, vol. 1, *The Early Oxford Schools*, Oxford: Clarendon Press, 401-33. (A lot of useful material about logic and logic teaching.)
- Libera, A. de** (1985-92) 'Bulletin d'histoire de la logique médiévale', *Revue des sciences philosophiques et théologiques* 69: 273-309, 71: 590-634, 76: 640-66. (A detailed, critical discussion of recent literature, with mini-introductions to a number of topics, including deontic logic and obligations.)
- Maierù, A.** (1994) *University Training in Medieval Europe*, trans. and ed. D.N. Pryds, Leiden: Brill. (This work, by a leading historian of logic, focuses particularly on how logic was taught, and its place in the curriculum, not just in universities, but in the *studia* of the religious orders.)
- Pinborg, J.** (1984) *Medieval Semantics: Selected Studies on Medieval Logic and Grammar*, London: Variorum. (A collection of seminal articles, published between 1969 and 1981.)
- Read, S.** (ed.) (1993) *Sophisms in Medieval Logic and Grammar: Acts of the Ninth European Symposium for Medieval Logic and Semantics, held at St Andrews, June 1990*, Dordrecht, Boston, MA, and London: Kluwer. (A very useful collection of articles.)
- Rijk, L.M. de** (1962-7) *Logica Modernorum: A Contribution to the History of Early Terminist Logic*, Assen: Van Gorcum, 2 vols in 3 parts. (A seminal collection of mainly twelfth-century texts with discussion. The source of much subsequent research, especially on supposition theory.)
- Spade, P.V.** (1988) *Lies, Language and Logic in the Late Middle Ages*, London: Variorum. (Articles on insolubles, Ockham's semantics, and obligations, by a leading historian of logic.)
- Stump, E.** (1989) *Dialectic and Its Place in the Development of Medieval Logic*, Ithaca, NY, and London: Cornell University Press. (A very readable collection of studies, and an excellent introduction to dialectic. Some of the material on consequences and obligations needs updating.)
- Tweedale, M.** (1988) 'Logic (i): From the Late Eleventh Century to the Time of Abelard', in P. Dronke (ed.) *A History of Twelfth-Century Western Philosophy*, Cambridge: Cambridge University Press, 196-226. (A good account of the early period, including Abelard.)

Logic, philosophy of

Philosophy of logic can be roughly characterized as those philosophical topics which have emerged either from the technical development of symbolic (mathematical) logic, or from the motivations that logicians have offered for their technical pursuits. In settling on a list of subjects to classify as philosophy of logic, therefore, there is a certain degree of arbitrariness, since the issues which emerge from the technical development of logic can equally well be assigned to such areas as semantics, philosophy of language, philosophy of mathematics, epistemology, and even ethics (see [Semantics](#); [Language, philosophy of](#); [Mathematics, foundations of](#)).

1 The impact of modal logic

In the broad area of mathematical logic, the biggest philosophical punch is packed by modal logic, including tense logic (see [Modal logic](#); [Modal logic, philosophical issues in](#); [Tense and temporal logic](#)). Modal logic has been important since Aristotle (see [Logic, ancient](#); [Logic in the 17th and 18th centuries](#); [Logic in the 19th century](#); [Logic in the early 20th century](#)) but has only been put on a rigorous footing in the second half of the twentieth century, by such figures as Hintikka, Kanger, Prior, and most especially Kripke (see [Semantics, possible worlds](#)). The most important philosophical outgrowth of this mathematical work is contained in Kripke's three lectures from January 1970 published as 'Naming and Necessity', in which Kripke draws out some ways in which possible worlds semantics is in tension with then-prevailing orthodoxies in the philosophy of language and mind. Some of Kripke's views have become new orthodoxies since (see [Essentialism](#); [Proper names](#); [Reference §§2-4](#); for related work by David Lewis, Robert Stalnaker, David Kaplan and others that uses the possible worlds framework, see [Counterfactual conditionals](#); [Demonstratives and indexicals](#); [Descriptions](#)).

To give some flavour of developments here, consider the familiar Fregean view that the relation of *reference* which holds between a name and its bearer is sustained by the relation of *presentation* which holds between the *sense* of the name and the bearer of the name: the name refers to such-and-such an object precisely because it expresses a sense which presents that object (see [Frege, G. §3](#); [Sense and reference](#)). When pressed for an explanation of what the senses of names are like, the natural Fregean response is to specify them, as Frege himself did in some cases, using definite descriptions (see [Descriptions](#)). So, for instance, the sense of the name 'Aristotle' might be 'the pupil of Plato who taught Alexander'. However, though it may well in fact have been Aristotle who taught Alexander, there are many ways things might have gone (many 'possible worlds') in which someone other than Aristotle is taught by Plato and teaches Alexander: suppose Aristotle had got the appointment but been killed in an accident before he could take it up, and had been replaced at Philip's insistence by another pupil of Plato. The description 'the pupil of Plato who taught Alexander' is therefore 'non-rigid', in Kripke's terminology. That is, it can pick out different individuals in different possible worlds, and in some worlds may pick out no one (Philip for some reason comes to distrust Platonic pedagogy and fails to conduct an equal opportunity search). But it is clear from the formal semantics for modal logic that there is conceptual 'room' for a category of expression which is 'rigid', in the sense that it picks out the same object in every possible world, or at least in every possible world where it picks out any object at all. So the formal semantics prompts the question whether names in natural language behave as if their reference is determined by a sense which presents different individuals at different worlds, or whether they behave as if they are rigid designators. With a series of brilliant examples Kripke demonstrates that names are rigid designators and therefore do not express reference-determining senses which are non-rigid (see [Proper names](#)).

The idea that a formal semantics for a kind of logic provides an account of a possible semantics for a category of natural-language expression, opening the door to debate on whether it is the right account or not, also captures some of the philosophical bearing of kinds of logic other than modal logic. Thus free logic shows how name-like expressions can function without standard existential commitment (see [Free logics, philosophical issues in](#)); intuitionistic logic and many-valued logic show how a language can have a compositional semantics even if its sentences are not used to make statements with verification-transcendent truth-conditions which always either obtain or fail to obtain (see [Compositionality](#); [Intuitionistic logic and antirealism](#); [Many-valued logics, philosophical issues in](#); [Presupposition](#)). And second-order logic offers a particular way of understanding the semantic import of a range of puzzling locutions, such as plural quantifiers (see [Second-order logic, philosophical issues in](#)). In all these cases the formal semantics for the logical system prompts debates about how well the

semantics carries over to natural language.

2 Logic and language

There is also a collection of long-established topics discussion of which can be much improved, in rigour at least, in the light of the development of modern logic. For example, a distinction between propositions (or statements, or sentential contexts) which are *de dicto* and propositions (and so on) which are *de re* originates in medieval philosophy. But only contemporary modal logic affords the tools for a precise characterization of this distinction, although it must be granted that the distinction remains a puzzle in *epistemic* contexts (see [De re/de dicto](#); [Descriptions §2](#); [Propositional attitude statements](#)). Other topics which can be classified in this way include [Essentialism](#), [Existence](#), [Identity](#), [Indicative conditionals](#), [Modal operators](#), [Quantifiers](#) and [Vagueness](#). Again, to give some of the flavour of this kind of work, consider the *de re/de dicto* contrast. There is an evident syntactic difference between ‘It is necessary that parents have children’ and ‘Parents are such that it is necessary that they have children’, but just because there is a syntactic difference, it does not follow that there is any interesting difference in meaning. But the difference can be brought out quite precisely in possible worlds semantics. To say that it is necessary that parents have children is to say that in every possible world, the people who are parents *in that world* have children in that world; and this is an obvious truth. On the other hand, to say that parents are such that it is necessary that they have children is to say that the people who are parents *in the actual world* are such that they have children in every possible world. This is clearly false, even putting aside contingency of existence of actual parents. For given anyone who is actually a parent, there is a way things could have gone - a possible world - in which that person is childless, hence not a parent (see [Quantifiers, substitutional and objectual](#); [Modal operators](#)).

When a formal semantics for a system of logic is applied to a fragment of natural language, a very precise account of the literal content of sentences in that fragment is given. But there may be aspects of the meanings of those sentences which are omitted. Philosophical views may then divide over whether the formal semantics has been shown to be wanting as an account of the semantics of the fragment, or whether instead the aspects of meaning not captured have been shown not to belong to literal content (see [Presupposition](#)). In the case of indicative conditionals, for instance, the formal semantics that is relevant is the simplest possible kind, namely, the truth-functional account of ‘if...then...’. According to this account, ‘If *p* then *q*’ is true if *p* is false or if *q* is true, regardless of the actual meanings of *p* and *q*. So in particular, any indicative conditional with a true consequent is true; examples would include ‘If lead floats in water then lead sinks in water’ and ‘If the solar system has nine planets then the Conservative Party lost the British elections in 1997’. Barring an astrological justification of the latter, both these conditionals look decidedly odd. But oddness is one thing, falsity another. The idea that such conditionals are false is based on the thought that if a conditional is true, then in establishing it in the most direct manner, non-redundant use has to be made of the antecedent. Spelling this out leads to relevance logic (see [Relevance logic and entailment](#); [Indicative conditionals](#)). On the other hand, if we say the conditionals are merely odd, we are led to some theory of communication to explain the oddness (see [Grice, H.P.](#); [Implicature](#)).

But we should not take away the impression that the traffic is all one way, from logic to language or from pure mathematics to pure philosophy. There is a two-way street here, with the above comments on conditionals representing a common phenomenon; that of a concern in the philosophy of language giving rise to a formal development which in turn feeds back into philosophy. For example, the idea that for a conditional to be true, the most direct way of establishing it must make non-redundant use of its antecedent seems clear enough on the face of it, but familiarity with logic of conditionals literature may well lead one to reconsider. This kind of dialectical interplay should continue to be a fruitful source of philosophical research for the foreseeable future.

See also: [Adverbs](#); [Ambiguity](#); [Anaphora](#); [Dummett, M.A.E.](#); [Fallacies](#); [Formal and informal logic](#); [Identity of indiscernibles](#); [Imperative logic](#); [Indirect discourse](#); [Intensional entities](#); [Intensionality](#); [Kripke, S.A.](#); [Linear logic](#); [Logic in China](#); [Logic in Islamic philosophy](#); [Logic in Japan](#); [Logical and mathematical terms, glossary of](#); [Logical constants](#); [Logical form](#); [Logical laws](#); [Mass terms](#); [Necessary truth and convention](#); [Ontological commitment](#); [Predication](#); [Prior, A.N.](#); [Propositions, sentences and statements](#); [Quantification and inference](#); [Quantifiers, generalized](#); [Questions](#); [Quine, W.V.](#); [Russell, B.A.W.](#); [Scope](#); [Type/token distinction](#); [Use/mention distinction and quotation](#); [Vagueness](#)

GRAEME FORBES

References and further reading

Hughes, R.I.G. (ed.) (1993) *A Philosophical Companion to First-Order Logic*, Indianapolis, IN: Hackett Publishing Company.(New and reprinted papers of varying levels of difficulty.)

Tomberlin, J.E. (ed.) (1994) *Logic and Language*, Philosophical Perspectives 8, Atascadero, CA: Ridgeview.(A collection of new papers on relevant topics.)

Logic, Renaissance

Renaissance logic is often identified with humanist logic, which is in some ways closer to rhetoric than to the study of formal argumentation. This is a mistake, for although changes did take place, a hard core of logical teaching remained the same throughout the medieval and Renaissance periods and into the eighteenth century. Logic was embedded in the educational system as the main study of beginning undergraduates, and although institutional changes had an effect on the presentation and use of logic texts, the study of valid arguments was always central.

There are two obvious differences between medieval texts and their sixteenth-century successors. The first is a new emphasis on following the order and material of Aristotle's *Organon*, with the consequent emphasis on the categorical syllogism as the central type of argument. Such medieval material as survived was strictly subordinated to this end; and even though the humanist logicians Agricola and Ramus had tried to ignore Aristotelian syllogistic and the doctrines propaedeutic to it (such as conversion and opposition), their omissions were rapidly remedied by subsequent textbook writers.

The second difference has to do with language and style. Medieval writers treated Latin as a technical, almost artificial language. They were deeply concerned with the effects that different word-orders and the addition of extra logical particles had on both meaning and reference, and they made heavy use of *sophismata*, deliberately constructed problematic or puzzling sentences. Although Latin remained the language of instruction, the approach of a Renaissance logician, whether humanist or Aristotelian, commentator or textbook writer, is totally different. *Sophismata* have completely disappeared, and so too has any attempt to treat Latin as a technical language in which different word-orders represent different logical structures. The propositions used for such operations as syllogistic conversion are presented in an already fully standardized form, and they are always relatively simple.

Why these changes came about is a difficult question. Humanism coexisted too long with medieval logic for humanism to be the sole explanation, and the fact that Renaissance logicians returned to the commentaries of Averroes and Aquinas on the *Organon* shows that mere revolt against anything medieval is not a sufficient explanation either. Changes in grammar teaching, changes in the relation of logic to the study of natural science, and changes in other parts of the university curriculum presumably have a good deal to do with the appearance of a new style of logic. In particular, the humanist emphasis on logic as a tool for analysing discourse focused attention on the use of logic in literature, history and biblical studies, and demanded a combination of simplicity and literary elegance, rather than any genuine formal innovation. There was no concurrent move to relate logic to the new developments in mathematics, and logic was not to be seen as a formal system linked with other formal systems until the nineteenth century.

1 The nature and divisions of logic

One of the features distinguishing sixteenth-century textbooks, whatever their orientation, from their medieval predecessors was their approach to the nature and divisions of logic. Medieval authors had asked whether logic is more properly a linguistic discipline directed towards spoken words (*scientia sermocinalis*) or a rational discipline directed towards concepts (*scientia rationalis*), and this question did not disappear (see [Logic, medieval §4](#)). However, its presentation was to some extent altered. The influence of the humanists, especially Agricola, gave general popularity to Cicero's definition (*Topics* II 6) of logic as an art of discourse (or *ratio disserendi*) (see [Cicero §2](#)), and this notion is also found in such scholastic authors as Toletus and Fonseca. At the same time, logic was frequently said to have as its subject either conceptual entities (beings of reason: *entia rationis*) (see [John of St Thomas §2](#); [Language, renaissance philosophy of §3](#)), or second intentions (see [Zabarella, J. §2](#)). Toletus denied a distinction between these two notions in this context, on the grounds that to talk about the kind of conceptual entity which has no existence independently of the mind is just to talk about those second-order logical concepts or intentions (genus, species, syllogism and so on) which we use to organize first-order concepts, those which apply directly to things in the world.

One issue which was the subject of lively debate concerned the very use of the word *logica* as opposed to *dialectica*. It was a medieval commonplace that *dialectica* could be used in two senses, a broad sense which equated dialectic with logic, and a narrow sense, in which dialectic was the kind of probable argumentation discussed in Aristotle's *Topics* (see §4) (see [Fonseca, P. da §2](#)). In the sixteenth century, Ramus argued in his

Scholae (Lectures) that the division of logic into demonstrative, dialectical and sophistical reasoning is baseless, and that all logic is dialectic (see §4). Zabarella, on the other hand, believed that ‘dialectic’ did name a distinct part of logic, and should be reserved for that part alone.

Another issue which reflects the impact of humanism concerned the way in which logic was divided. Perhaps the most important division was that into invention and judgment, invention being the part of logic which finds the subject matter for argumentation, and judgment being the part which organizes the subject matter. This division came principally from Cicero (*Topics* II 6) and was well known to medieval logicians, before it was further popularized by Agricola, Melanchthon and Ramus. A mark of Ramist textbooks was that they began with invention, but Melanchthon, followed by various authors including John Seton (see §5) put judgment first. Another way of organizing which was to become especially popular in the seventeenth century (see John of St Thomas §2; Logic in the 17th and 18th centuries §1) was a division in accordance with the three acts of mind, the apprehension of simples (terms), composition and division (propositions), and reasoning (arguments). Like the division into invention and judgment, this division has a long history.

2 The medieval inheritance

Although Aristotle served as a focal point for the study of logic during the Middle Ages and Renaissance (see §3), one of the main features of medieval logic was the production of independent texts on particular topics, some of which had no reference to Aristotelian logic. The most important of these texts were the so-called *parva logicalia*, or treatises dealing with the properties of terms, including their reference in various contexts. Here we find tracts on supposition, on relative terms, on ampliation, appellation, restriction and distribution (see Logic, medieval §7). These core treatises were supplemented in various ways, particularly by treatises about exponible terms, those logical particles such as ‘except’ and ‘only’ whose presence calls for the analysis of an apparently simple categorical proposition into several conjoined propositions.

A second major group of treatises consisted of the ‘three tracts of the moderns’, which concern propositions and the relations between them. Here we find treatises on consequences, obligations and insolubles. Treatises on consequences covered all types of argumentation, beginning with a general discussion of what constitutes a formally valid inference, and including a good deal of propositional logic. The syllogism often appeared as a special example of one kind of argumentation. Treatises on obligations dealt with the rules to be followed in a certain kind of disputation which was specifically designed to test the logical skills of undergraduates, and which deliberately confined itself to exploring the logical consequences of affirming an often bizarre falsehood. Treatises on insolubles dealt with semantic paradoxes, such as the standard liar paradox, ‘What I am now saying to you is false’, and they explored in some depth the semantic presuppositions of language (see Semantic paradoxes and theories of truth).

During the fifteenth century much of this medieval material was used throughout Europe, and Italy was particularly enthusiastic about the products of fourteenth-century Oxford logicians, such as Heytesbury and Ralph Strode. *Summulae* or general textbooks from the earlier period were also used. The best known is by Peter of Spain, and it gives a fairly complete outline of Aristotelian logic as well as the *parva logicalia*. However, it would be a mistake to think that only Peter of Spain was used, for in some places (such as England) he was never used, and in other places only parts of his work were used. Buridan (§2), Albert of Saxony and Paul of Venice were the authors of other popular textbooks. Indeed, the first medieval logic text known to have been printed is Paul’s *Logica parva*, printed in 1472 (and still being reprinted in Venice as late as 1614).

At the beginning of the sixteenth century, particularly in Paris, there was a strong revival of the medieval tradition. Separate treatises on supposition theory, exponibles, consequences, insolubles and obligations continued to be written (see Major, J. §2), and to these were added new treatises on such matters as terms and opposed propositions. At the same time, new commentaries on Peter of Spain were being published in Paris, Germany and Spain. The 1529 edition of Domingo de Soto’s *Summulae* included commentary on the first and fourth books of Peter of Spain along with treatises dealing with terms and supposition theory, exponibles, insolubles, and obligations (see Soto, D. de).

Outside Spain, this activity came to an abrupt halt around 1530. No new independent treatises on Peter of Spain were written; the publication of works written during the first three decades of the sixteenth century ceased, and

the publication of such medieval authors as Buridan and Heytesbury also ceased. Only in Spain and its colonies did authors continue to produce commentaries on Peter of Spain, such as that by Alonso de la Vera Cruz, first published in Mexico in 1554, or that by Tomás Mercado, published in Seville in 1571 (see [Latin America, colonial thought in §3](#); [Mexico, philosophy in §1](#)).

By the end of the sixteenth century those parts of medieval material which could not easily be integrated with Aristotelian logic simply disappeared, at best being the subject of stray and often unflattering references. Insolubles were no longer discussed, obligations were replaced by a discussion of ordinary disputations in which truth was the main issue, and many of the topics connected with supposition theory were dropped. But exponibles, consequences and the central part of supposition theory remained in truncated and simplified form (see §6).

3 Aristotle and Aristotelianism

At least in principle, Aristotle's *Organon* played a central role throughout the medieval and Renaissance period (see [Aristotle §4](#)), though not all parts were studied with equal fervour. For instance, the *Topics* was sometimes omitted altogether, and sometimes only certain of its books were assigned. The text of the *Organon* was accompanied by explanation and commentary, and the invention of printing brought the wide dissemination of some of the medieval commentaries on Aristotle's logical works, including those by Walter [Burley](#) and [Paul of Venice](#). Later there was a return to commentaries by the great theologians of the thirteenth and early fourteenth centuries. Thus, the commentaries of Aquinas on *De interpretatione* and *Posterior Analytics* reached a peak of popularity in the mid-sixteenth century, and the commentaries by or attributed to [Albert the Great \(§2\)](#), [Giles of Rome](#) and John Duns Scotus were also published (see [Aristotelianism, medieval §5](#)). The great Giunta edition of Aristotle with the commentaries of Averroes (1550-2) should also be mentioned.

The impact of humanist classical studies on Renaissance logic is particularly important in two respects: first, the edition and translation of Greek commentators on Aristotle, only some of whom had been known during the Middle Ages; and second, the edition and translation of Aristotle himself. During the fifteenth century the Byzantine humanist and philosopher Johannes Argyropoulos (c.1415-87) had produced new Latin translations, but as the Greek text became better known, more and more dissatisfaction was felt with former translations, including those of Argyropoulos. A good example of the new Greek-based texts is the *Organon* edition of Giulio Pace first published in 1584. In it we find the Greek text side-by-side with a new translation designed not only to read well but also to capture the philosophical significance of Aristotle's words. In the margins there is a commentary dealing with difficult points both of theory and of translation. However, the old translations did not disappear, because they provided much of the standard vocabulary in philosophy, and the work of surviving medieval commentators was keyed to them (see [Humanism, Renaissance](#)).

So far as content is concerned, there were few developments within Aristotelian logic. There was considerable discussion of scientific method (see §7), and there was also some lively debate about the acceptability of the fourth figure of the syllogism, particularly inspired by [Zabarella](#). Otherwise, Aristotelianism's main achievement was its successful stand against attempts such as those by Lorenzo [Valla \(§4\)](#) and Petrus [Ramus](#) to replace basic Aristotelian notions. This achievement was mainly due to the felt need for formal content in logic teaching (see §5) (see [Aristotelianism, Renaissance §1](#)).

4 Humanist logic and the topics

To understand humanist logic, one must consider the relations between dialectic, the topics, and rhetoric. 'Dialectic' is a term with various meanings. For those who did not use it as a mere synonym for logic, at least three approaches were possible. One could present dialectic as the study of debate or disputation; one could present dialectic as the study of the topics (roughly speaking, headings under which material could be gathered); or one could present dialectic as the study of probable argumentation.

The last two approaches are closely related to the concerns of rhetoric, and there is a tendency in the literature to speak as if humanist logic was the result of subordinating logic to rhetoric. However, while it is true that rhetoric texts featured the invention or finding of topics, and that rhetoric, as the art of persuasion, could be construed as dealing with arguments about probabilities, the three standard works about topics had always been firmly part of the logic syllabus. These were Aristotle's *Topics*, together with Boethius' two works, *De topicis differentiis* (*On*

Topical Differences) and *In Ciceronis Topica* (*On Cicero's Topics*) (see [Boethius, A.M.S. §3](#)). Boethius' first work was more important in the medieval period, though in the fourteenth century it was largely replaced by the account of topics given by Peter of Spain in his *Summulae*. In the sixteenth century Boethius' second work became popular again, as did Cicero's original text. Agricola himself was well-acquainted with both Aristotle and Boethius.

The notion that dialectic is concerned with probable argumentation comes from Aristotle's *Topics*, which opens with a distinction between demonstrative reasoning, said to deal with certainties, and dialectical reasoning, said to deal with probabilities. However, there are two ways of reading the distinction. One way involves two types of logic, a logic of formally valid deductive arguments, and a logic of informal arguments or strategies. The other way of reading Aristotle's distinction is to assume that 'argument' or 'reasoning' in the *Topics* refers only to the syllogism as described in the *Prior Analytics*, namely a particular kind of formally valid deductive structure. This second reading led medieval logicians to interpret Aristotle's distinction between demonstrative reasoning and dialectical reasoning as an epistemic one. Thus, demonstrative reasoning involves formal syllogisms with premises that are certainly true, and dialectical reasoning involves the same formal syllogisms with premises that are only probably true. It was this epistemic reading that led Ramus to argue that dialectic and logic are the same, since the same patterns of argumentation appear in both. In Agricola, on the other hand, we find a somewhat greater awareness of the possibility of a properly non-formal logic.

The general failure of humanist logicians, including Agricola, to pursue dialectic as an informal logic of probabilities was closely associated with their treatment of the other aspect of dialectic, dialectic as a topics-logic, or means of finding the material for arguments. In Boethius, the notion of a topic had covered two things. First, a topic was a maxim (or 'maximal proposition'), a self-evidently true universal generalization which could either be inserted into an argument as a premise or which could be appealed to as providing a warrant for an argument. Some of these maxims were themselves turned into formal deductive arguments by medieval logicians, who had a tendency to absorb invention into judgment, but others were not easily formalized. As such, they provide the perfect nucleus for a developed informal logic. The second sense of topic for Boethius was the topic as the *differentia* of a maximal proposition, that is, the characteristic that enables us to classify maximal propositions into groups. To list the topics in this second sense is simply to list the headings under which material can be gathered: 'definition', 'genus', 'cause', 'opposite', 'similar' and so on.

What we find in [Agricola \(§2\)](#) is a deliberate rejection of the maxims, and of material that might belong to judgment rather than invention. He argued that the purpose of any maxim is simply to present a necessary argument, and that they are quite inappropriate for all the cases in which we are dealing with probabilities. We should note here that Agricola is clearly talking not just of the epistemological status of the premises of an argument, but of a probabilistic and informal relationship between premises and conclusion. He also claimed that maxims were too restrictive, for they represent an attempt to force a wide range of material into a narrow compass; and that they are of no use to someone who really understands the nature of topics. What we are left with, then, is the topics solely as headings under which material can be organized. Their most obvious link with argumentation, both formal and informal, has been broken.

In the hands of Agricola's later follower, Petrus [Ramus \(§2\)](#), we find the topics presented in an Agricolan way, without maxims, but with a new theoretical foundation. They are not merely useful headings for gathering material, but they represent the mind's natural organization of data. As such, topics function as categories; and, indeed, Ramus emphasized that his topics are more useful and more natural than Aristotle's ten categories (see [Categories §1](#)). If we are asked to discuss war, peace, or the state, it is no use to think in terms of the category of substance. Instead we must appeal to such topics as cause, event and opposites.

The result of these reforms was a subordination of rhetoric to logic (rather than the reverse), in three respects. First, the topics were now to be confined to dialectic, and rhetorical topics, which had covered material of a more particular kind, were to be reduced to dialectical topics. Second, rhetoric's *dispositio* is introduced into the dialectic text through the discussion of method as a way of ordering discourse. Third, we find that rhetoric, at least in the eyes of the humanist logician, was to be confined to the business of ornamenting discourse, and discussing only those persuasive devices which did not involve argumentation.

5 Humanist and Ramist textbooks

The needs of the university curriculum meant that humanist insights and changes were only partially reflected in logic teaching. Thus Agricola's seminal *De inventione dialectica* (see [Agricola §2](#)) was often used in conjunction with a little Aristotelian handbook by [George of Trebizond](#). The very popular textbooks of [Melanchthon](#) provide another good example. He enjoyed the Agricolan emphasis on clarity of style and the use of literary allusions; he accepted the importance of the topics and the part of logic called invention; and he discussed method. At the same time, he remained a convinced Aristotelian, who believed that students needed to be taught some formal logic. The formal techniques he used were those of syllogistic, and his work included a discussion of the other standard Aristotelian logical subjects, including the categories and the square of opposition for propositions.

Another example from the first part of the sixteenth century is John Seton's *Dialectica*, first published in England in 1545, and later to appear with annotations by Peter Carter. Seton explained in his introduction that he had written the work because of the absence of a suitable text for the instruction of the young. Aristotle was too difficult, Agricola had deliberately restricted himself to the subject of invention, and Melanchthon's style was not suited to elementary teaching. He makes frequent references to Agricola, as well as to [Cicero](#), Quintilian and [Erasmus](#). He included the bulk of traditional Aristotelian logic from the categories to the syllogism, which he described as the most important part of judgment, and he also retained the strictly medieval doctrine of the supposition of terms, though in a considerably truncated form.

In the second part of the century the most important textbook writer was Petrus [Ramus](#) ([§2](#)), the most notorious logician of the period. He is known both for his attacks on Aristotle and for the simplified logic presented in his *Dialectique* of 1555, and in a variety of Latin versions. The *Dialectique* had two parts. The first, on invention, covered the topics, and the second, on judgment, presented a deliberately simplified version of the syllogism followed by an account of method as a means of ordering in the arts and sciences. No reference was made to such standard material as the categories, the square of opposition, conversion, demonstration and fallacies. On the other hand, the work is rich with quotations from the poetry and prose of classical authors, which must have strengthened the impression among students that logic was both easy and fun.

The book had a remarkable publishing history, but it is not clear how much impact it actually had on the university scene. In England, Ramus certainly enjoyed some popularity at Cambridge but was less well received at Oxford. Even in Germany, where 151 editions appeared, university teachers of logic soon found serious deficiencies in Ramus' book, and in the 1590s a new school of textbook writers known as the Philippo-Ramists appeared in Germany. Their aim was to combine what was best in Ramus with what was best in the more Aristotelian work of Philipp Melanchthon. As a result, they tended to restore those parts of Aristotelian logic that Ramus had deliberately omitted, and the syllogism was once more presented as central to formal logic (see [Logic in the 17th and 18th centuries §1](#)).

6 Late scholastic textbooks

If England and Germany saw the production of slim textbooks exemplifying the marriage between humanism and a simplified Aristotelianism, Spain and Italy were the scene of a much more substantial textbook production. Two important texts, both recommended by the Jesuit *Ratio Studiorum* of 1599, and thus possibly studied by Descartes at La Flèche, were the *Introductio in dialecticam Aristotelis* (*Introduction to the Dialectic of Aristotle*) of Franciscus [Toletus](#) ([§3](#)) and the *Institutionum dialecticarum* (*Dialectical Instructions*) of Pedro da [Fonseca](#) ([§2](#)). Toletus' book was first published in Rome in 1560, and the last of its eighteen editions appeared in Milan in 1621. Fonseca's work was first published in Lisbon in 1564, and the last of its fifty-two editions appeared in Lyons in 1625. The works were fairly similar in content. The Organon is very much central but medieval material is used to supplement Aristotle in various places. Exponibles are discussed, albeit in a simplified form; some non-syllogistic consequences dealing with truth and modality are listed, and material about supposition and related doctrines is presented as an aid to understanding fallacies.

The most important English text of this type was Robert Sanderson's *Logicae Artis Compendium* which appeared at Oxford in 1615 and was to remain a standard text there until well into the eighteenth century. This, too, contains a full account of strictly Aristotelian logic, together with brief discussions of the medieval doctrines of supposition, exponibles and consequences, and a full discussion of method as a way of ordering discourse.

7 The discussion of method

A particular feature of textbooks, especially in the second part of the sixteenth century, was the discussion of method. First, and most important, is the study of scientific method, which is particularly characteristic of Italian universities. These had always been distinguished from Northern European universities by their strong emphasis on law and medicine, combined with a relatively slight emphasis on theology. The faculties of arts at such places as Padua provided studies leading to a degree in arts and medicine, and as a result the main emphasis was placed on logic and natural philosophy, as propaedeutic to medicine. In the fifteenth century works stemming from Oxford and Paris on such topics as the intension and remission of forms were particularly important (see [Natural philosophy, medieval §6](#)), but after about 1520 these works fell into sudden oblivion. By the mid-sixteenth century the focus of attention was on the Greek commentators and on Averroes (see [Ibn Rushd](#)), especially his *Physics* commentary. The main Aristotelian text studied was the *Posterior Analytics*. From Cajetan of Thiene (Gaetano da Thiene) (1387-1465) to [Nifo](#) there was a gradual refinement of writings on demonstration and scientific method which culminated in the work of Jacopo [Zabarella §4](#). His discussions of the methods of resolution, composition, and the so-called *regressus* (a method for uniting resolution and composition) were summarized in many later textbooks, often in uneasy alliance with two other types of method.

A second notion of method presents it as a way of organizing discourse, a method fully discussed by [Agricola §2](#). Agricola showed how argumentative and expository passages from classical literature could be analysed, a method which brought the logical text very close to the new humanist curriculum of many institutions (see [Humanism, Renaissance §7](#)). Agricola's approach was picked up and further popularized by [Melanchthon §2](#).

A third type of method is that of [Ramus §3](#), which is particularly directed towards the presentation of a subject, and best known for its presentation through diagrammed dichotomies. This too could be and was adapted to the analysis of literary and biblical texts.

See also: [Logic, ancient](#); [Aristotelianism, medieval](#); [Aristotelianism, Renaissance](#); [Aristotle](#); [Humanism, Renaissance](#); [Language, medieval theories of](#); [Language, Renaissance philosophy of](#); [Logic, medieval](#)

E.J. ASHWORTH

References and further reading

- Ashworth, E.J.** (1974) *Language and Logic in the Post-Medieval Period*, Dordrecht and Boston, MA: Reidel. (A comprehensive account, thematically arranged, and making use of the techniques of modern symbolic logic.)
- Ashworth, E.J.** (1978) *The Tradition of Medieval Logic and Speculative Grammar from Anselm to the End of the Seventeenth Century: A Bibliography from 1836 Onwards*, Toronto, Ont.: Pontifical Institute of Mediaeval Studies. (Annotated bibliography with full indexes.)
- Ashworth, E.J.** (1985) *Studies in Post-Medieval Semantics*, London: Variorum. (A collection including studies of supposition theory, insolubles and exponibles.)
- Ashworth, E.J.** (1988) 'Traditional Logic', in *The Cambridge History of Renaissance Philosophy*, ed. C.B. Schmitt, Q. Skinner and E. Kessler, Cambridge: Cambridge University Press, 143-72. (A useful general account with plenty of references.)
- Jardine, L.** (1988) 'Humanistic Logic', in *The Cambridge History of Renaissance Philosophy*, ed. C.B. Schmitt, Q. Skinner and E. Kessler, Cambridge: Cambridge University Press, 173-98. (A useful general account, though controversial on Valla; shows how humanist logic was used in textual exegesis.)
- Lohr, C.H.** (1988) *Latin Aristotle Commentaries II. Renaissance Authors*, Florence: Olschki. (Alphabetically organized list of Renaissance Aristotelians, with biographies and bibliographies of primary and secondary sources.)
- Mack, P.** (1993) *Renaissance Argument: Valla and Agricola in the Traditions of Rhetoric and Dialectic*, Leiden, New York and Cologne: Brill. (Excellent account of humanist logic, including Melanchthon and Ramus.)
- Risse, W.** (1964) *Die Logik der Neuzeit. I Band. 1500-1640 (The Logic of the Modern Period)*, Stuttgart and Bad Cannstatt: Frommann. (A dense, scholarly account, without much awareness of the medieval background, and with no awareness of twentieth-century logical developments.)
- Risse, W.** (1965) *Bibliographia Logica. Verzeichnis der Druckschriften zur Logik mit Angabe ihrer Fundorte. Band I. 1472-1800 (Bibliography of Logic: List of Printed Works on Logic with Details of their Location)*, Hildesheim: Olms. (The bibliography organized chronologically but it is fully indexed.)
- Sanderson, R.** (1618) *Logicae Artis Compendium*, ed. E.J. Ashworth, Bologna: Editrice CLUEB. (This book

contains a facsimile of the 1618 edition of Sanderson's work (referred to in §6), together with indexes and a lengthy introduction which gives full information about English logic in the sixteenth century, including a discussion of John Seton (referred to in §5).)

Vasoli, C. (1968) *La dialettica e la retorica dell'Umanesimo. 'Invenzione' e 'Metodo' nella cultura del XV e XVI secolo* (Humanist Dialectic and Rhetoric: 'Invention' and 'Method' in the Culture of the 15th and 16th Centuries), Milan: Feltrinelli. (Dated but classic account of humanist logic.)

Wallace, W.A. (1995) 'Circularity and the Paduan *Regressus*: From Pietro d'Abano to Galileo Galilei', *Vivarium* 33: 76-97. (Outline of Paduan discussions of scientific method.)

Logical constants

A fundamental problem in the philosophy of logic is to characterize the concepts of ‘logical consequence’ and ‘logical truth’ in such a way as to explain what is semantically, metaphysically or epistemologically distinctive about them. One traditionally says that a sentence p is a logical consequence of a set S of sentences in a language L if and only if (1) the truth of the sentences of S in L guarantees the truth of p and (2) this guarantee is due to the ‘logical form’ of the sentences of S and the sentence p . A sentence is said to be logically true if its truth is guaranteed by its logical form (for example, ‘2 is even or 2 is not even’). There are three problems presented by this picture: to explicate the notion of logical form or structure; to explain how the logical forms of sentences give rise to the fact that the truth of certain sentences guarantees the truth of others; and to explain what such a guarantee consists in.

The logical form of a sentence may be exhibited by replacing nonlogical expressions with a schematic letter. Two sentences have the same logical form when they can be mapped onto the same schema using this procedure (‘2 is even or 2 is not even’ and ‘3 is prime or 3 is not prime’ have the same logical form: ‘ p or not- p ’). If a sentence is logically true then each sentence sharing its logical form is true. Any characterization of logical consequence, then, presupposes a conception of logical form, which in turn assumes a prior demarcation of the logical constants. Such a demarcation yields an answer to the first problem above; the goal is to generate the demarcation in such a way as to enable a solution of the remaining two.

Approaches to the characterization of logical constants and logical consequence are affected by developments in mathematical logic. One way of viewing logical constanthood is as a semantic property; a property that an expression possesses by virtue of the sort of contribution it makes to determining the truth conditions of sentences containing it. Another way is proof-theoretical: appealing to aspects of cognitive or operational role as the defining characteristics of logical expressions. Broadly, proof-theoretic accounts go naturally with the conception of logic as a theory of formal deductive inference; model-theoretic accounts complement a conception of logic as an instrument for the characterization of structure.

1 The Tarskian paradigm

Suppose we are given a theory of truth for a language L , a theory which explains how the truth-conditions of sentences of L are determined by their construction from the semantic properties of a set of primitive expressions. A semantic characterization of logical consequence and logical truth for L appeals to the idea of an ‘interpretation’ of L , that is, roughly, an assignment of semantic properties of the appropriate sort to the nonlogical (or ‘lexical’) primitive expressions of L , properties which, once assigned, determine a (unique) truth-value for each sentence in L . If p is a sentence and S a set of sentences of L , then p is said to be a logical consequence of S if p is true on each interpretation of L rendering every sentence of S true, and p is logically true if it is true on each interpretation of L .

The technical implementation of this idea for elementary logic goes back to Löwenheim (1915), but its full significance was appreciated only in the 1930s; it receives an important exposition in Tarski’s 1936 paper ‘Über den Begriff der logischen Folgerung’ (‘On the Concept of Logical Consequence’). Let L be an elementary first-order language. The logical constants of L are given by enumeration as the four connectives ‘ $\neg$ ’ (negation), ‘ $\&$ ’ (conjunction), ‘ $\vee$ ’ (disjunction), ‘ $\rightarrow$ ’ (material conditional) and the two quantifiers ‘ $\forall$ ’ (universal quantifier) and ‘ $\exists$ ’ (existential quantifier). In some cases, the identity symbol ‘ $=$ ’ is also distinguished as a logical constant. The formulas of L arise in the following three ways: (1) by applying any primitive n -place predicate symbol R to n terms $t_1, \dots, t_n$ to obtain a formula $R(t_1, \dots, t_n)$; (2) by applying connectives to given formulas; and (3) where $x_0, x_1, \dots$ is a fixed set of individual variables, by applying the quantifiers ‘ $\forall$ ’ and ‘ $\exists$ ’ to a variable x_n and formula p to give formulas $(\forall x_n)p$ and $(\exists x_n)p$. An interpretation of L consists, first, of a non-empty set D , its domain or universe; second, of an assignment of relations and functions of the appropriate type over D to the predicate and function symbols of L ; and third, of an assignment of objects in D to each individual constant (or name) in L . If the identity predicate is treated as a logical constant, it is understood that the interpretation assigns to it the identity relation over D , that is, the set $\{\langle a, a \rangle : a \in D\}$.

If J is an interpretation of L , the meanings of the logical expressions determine, for each formula of L , when the

formula is true on J relative to an assignment of objects in D to the individual variables of L . Such an assignment may be represented by a sequence over D , that is, a function mapping the natural numbers $\{0, 1, 2, \dots\}$ into D . A sequence s then expresses the assignment to each variable x_n of the object $s(n)$ from D . A formula which is true on J with respect to an assignment s is said to be ‘satisfied’ by s in J . The sequences which satisfy a given formula on J are determined by those which satisfy its component formulas (subformulas). Thus, for example, a sequence satisfies the conjunction of two formulas on J when it satisfies both conjuncts, and satisfies the negation of a formula when it fails to satisfy the formula. Finally, if $q = (\exists x_n)p$, a sequence s satisfies q on J if and only if there exists an assignment s' differing from s only in the value assigned to the variable x_n such that s' satisfies p in J . This expresses the idea that q should be true on an assignment if it is possible to vary the object assigned to the quantified variable in such a way as to make p true. An analogous clause may be given for the universal quantifier. The interpretation J is said to be a ‘model’ of a set S of formulas with respect to an assignment if each formula in the set is true in J with respect to that assignment; S is said to be ‘consistent’ if it has a model. Tarski then says that a formula p is a logical consequence of a set S of formulas if and only if for any interpretation J of L and assignment s in J , if J is a model of S with respect to s , then p is true in J with respect to s ; equivalently, p is a logical consequence of S if the result of adding $\neg p$ to S is inconsistent. A formula is logically true if it is true in each interpretation of L with respect to each assignment; and two formulas are said to be logically equivalent when they are true in just the same interpretations relative to just the same assignments (see [Tarski, A.; Model theory; Tarski’s definition of truth](#)).

2 Extensions of elementary logic

It is fair to say that the account of logical consequence and logical truth we have just sketched for elementary logic has set the standard for semantic accounts of these notions in other contexts. Typical of these are extensions of elementary logic that arise by treating additional quantifiers as logical constants.

A first-order quantifier Q may be regarded as describing a relation between a universe D (any non-empty set) and the subsets E of D such that the sentence $(Qx)(Ax)$ is true on an interpretation that assigns E to $A(x)$ over the universe D . Consider, for example, the quantifier symbol Q_0 , read ‘for infinitely many’. The sentence $(Q_0x)A(x)$ holds in an interpretation J if $A(x)$ holds in J relative to infinitely many values for the variable x . In effect, then, Q_0 describes the relation between a universe and its infinite subsets. Similarly, the quantifier Q_1 describes the relation between a universe and its uncountable subsets (if any), where a set is called ‘countable’ if it can be put into a one-to-one correspondence with a subset of the natural numbers. Q_0 and Q_1 are examples of ‘generalized quantifiers’ (see [Quantifiers, generalized](#)).

Elementary logic is characterized by three properties which are focal points of any evaluation of its extensions. The ‘completeness theorem’ for elementary logic states that if a formula p is a logical consequence of a set S of formulas in elementary logic, then p is in fact provable from S within the deductive systems for elementary logic usually considered. Combined with the soundness theorem, which states the converse property, this result has the effect of saying that the logical consequence relation for an elementary language can be represented syntactically: the consequences of any set are precisely the formulas that stand to that set in a syntactically characterized relation of deducibility. The ‘compactness property’ holds for a relation of logical consequence just in case a formula is a consequence of an infinite set S of formulas only if it is a consequence of some finite subset of S . Finally, the (downward) Löwenheim-Skolem theorem for elementary logic states that if a set of formulas has a model, then it has a model with a countable universe. The quantifiers Q_0 and Q_1 properly extend elementary logic in rather different ways: the consequence relation that arises by adding Q_0 to elementary logic has the Löwenheim-Skolem property but not the compactness or completeness properties; while the corresponding relation derived from Q_1 has both the completeness and compactness properties, but not the Löwenheim-Skolem property (see [Keisler 1971](#)). Other extensions of elementary logic have neither property. An example is second-order logic, which extends elementary logic by allowing quantification into predicate places (see [Löwenheim-Skolem theorems and nonstandard models; Second-order logic, philosophical issues in](#)).

3 Notions of invariance

In a model-theoretic definition of logical consequence, the logical constants are given simply by a list. The choice of one such list over another may seem, to some extent, ungrounded. Thus, Tarski wrote in [1936](#) that ‘no objective

grounds are known to me which would permit us to draw a sharp boundary between the two groups of terms', that is, to solve the demarcation problem. However, in later years Tarski was led to a less sceptical attitude towards the demarcation problem (see [Tarski 1966](#)). The solution he suggested consisted of a very general condition - it is satisfied, in fact, by all of the generalized quantifiers usually considered - but one strong enough to exclude many of the expressions which are generally regarded as nonlogical. Let us formulate the requirement for the case of a quantifier symbol Q that applies to a formula $A(x)$ and variable x to yield a sentence $(Qx)(Ax)$. (The generalization of this constraint to other classes of expressions is straightforward and may be found in the literature - see McCarthy ([1981: §3](#)) and Sher ([1991: ch. 4](#).) The meaning of Q must determine, for each chosen universe D and each interpretation of $A(x)$ over D , a truth-value for the sentence $(Qx)(Ax)$ over D . Now consider any one-to-one function mapping D on to an alternative universe D^* . Such a function fixes an interpretation for $A(x)$ over D^* , consisting precisely of the images under the function of the items in the given interpretation of $A(x)$ over D . The requirement is then that the two interpretations assign the same truth-value to $(Qx)(Ax)$. Such a function is said to constitute an 'isomorphism' of the one interpretation with the other; and so a succinct statement of the indicated condition is that the behaviour of a logical constant must be 'invariant under isomorphism of interpretations'. Tarski's suggestion is that the boundary between logical and nonlogical notions coincides with the boundary between the expressions which satisfy the present invariance condition and those which do not.

What is the significance of the invariance requirement? Interpretations are 'isomorphic' when they share the same structure, so the invariance condition says roughly that the semantic result of applying a logical constant should depend only on the structure of the interpretation of the expressions to which it is applied, independently of properties and relations specific to objects in any particular interpretation. This requirement may be used to confer a clear sense on one very traditional thesis about logical notions, namely that they are 'topic-neutral' concepts. Tarski noticed that this sort of topic-neutrality is related to a unifying idea behind certain generalizations of the notion of a 'geometry', an idea developed by the German mathematician Felix Klein ([1872](#)). Klein viewed a space S as being given by a set with a certain geometric structure. Consider a collection G of one-to-one transformations mapping this set onto itself. Geometric structures can in many cases be described by families of such transformations. For example, consider the collection of all transformations that can be defined in terms of rotations and linear motions (translations). Then the properties of (plane) Euclidean geometry are characterized by the fact that they are 'invariant under transformations of this type', in the sense that any such transformation maps a set S of points on to a set congruent to S and, conversely, any set congruent to S can be obtained from S by a transformation of this type. Similarly, topology, a generalized geometry, may be viewed as the study of those properties which are preserved by one-to-one transformations mapping a space 'continuously' (intuitively, without breaking or tearing it) onto itself. Tarski's idea was to use the limiting case of this situation to characterize logic: logic is the study of those properties of a 'space' (that is, a set or universe) which are preserved by *any* one-to-one transformation mapping that space onto itself or, more generally, onto some other set. The indicated invariance requirement for Q may be seen as the special case of Tarski's constraint where we focus on relations of a universe and its subsets of the sort described by quantifiers.

The invariance condition is generally regarded as necessary for logical status. The question we must now consider is whether it is also sufficient. One difficulty faced by an affirmative answer concerns the necessitation requirement: that it is not possible - in a strong, uniquely logical sense of 'possible' - that each member of a set of sentences is true and one of its logical consequences is false. The difficulty is that there appear to be expressions which are counted as logical by the invariance criterion which lead to consequence relations violating this requirement. For example, let P represent a fixed contingently false sentence of English, for example, 'The French Revolution failed'. Consider a quantifier symbol Q interpreted as follows: in any interpretation J , $(Qx)(Ax)$ holds in J if P and $A(x)$ holds for some member of J , or not- P and $A(x)$ fails for each member of J . Since P is in fact false, if Q were a logical constant then $(Qx)(Ax)$ would be logically equivalent to $(\forall x)\neg A(x)$. This is a violation of the necessitation requirement, since in a logically possible world in which P is true, $(Qx)(x = x)$ is true, but $(\forall x)\neg(x = x)$ is false.

However, it may be that the lesson to be learned from examples such as the one just considered is not that the invariance constraint fails to underwrite the necessitation requirement, but that the invariance requirement has been understood too narrowly (see [Sher 1991](#); [McCarthy 1987](#)). The example would not arise if that requirement were

understood ‘rigidly’, that is, in such a way as to apply to the interpretation of Q ‘across all logically possible worlds’. For consider a world w_1 in which P is true containing the same objects as a world w_2 in which P is false. Then the transformation that associates each object with itself will map the interpretation of the formula $x = x$ in w_1 on to its interpretation in w_2 (both are the whole universe), but the sentence $(Qx)(x = x)$ is true in w_1 and false in w_2 , which shows that the invariance condition for Q breaks down if that condition is applied to the interpretation of Q across all logically possible worlds. Let us say that an expression satisfying this stronger invariance condition is ‘rigidly invariant’. The suggestion is then that rigid invariance be taken as the defining property of the logical constants. The problem created by this proposal lies in a threat of circularity: we would require a characterization of the notion of logical possibility, sufficient for a demarcation of the logically possible worlds, not dependent upon a prior demarcation of the logical constants.

4 Syntactic characterizations of logical notions

An alternative approach to the demarcation problem locates the distinctive feature of the logical constants in the role that they play in deductive arguments. On this view, a logical expression is one whose meaning can be characterized by a set of rules of inference of a certain sort. These rules may be formulated in different ways; for example, in terms of a sequent calculus, or a natural deduction system (see [Natural deduction, tableau and sequent systems](#)). Each formulation will incorporate in some way a distinction between rules that license the introduction of a logical expression in an inference (introduction rules), and those which describe the conditions under which it can be eliminated (elimination rules). Thus, for example, the introduction condition for conjunction may be given by the rule

$A, B / A \ \& \ B,$

and its elimination conditions by the rules

$A \ \& \ B / A, \quad A \ \& \ B / B.$

A rule of the form

$X_1, \dots, X_n / Z$

is to be read as an inference-licence, saying that if $X_1, \dots, X_n$ have been inferred, then it is permissible to infer Z . The existential quantifier may be similarly characterized by the rules

$A(t) / (\exists x)A(x), \quad (\exists x)A(x) / A(c),$

where in the first rule t is any term and in the second c is a constant subject to the restriction that it not appear previously in the inference in question (the ‘non-occurrence restriction’). Analogous rules characterize the other notions of classical quantification theory.

The most straightforward reading of the syntactic view is that an expression counts as logical if and only if it is possible to write a set of introduction and elimination rules for it. Arthur Prior criticized this idea by means of a simple counterexample (1960). In the absence of substantive restrictions governing the interaction of introduction and elimination rules, it would seem that we could characterize a connective ‘tonk’ by means of the following pair of rules:

$A / A \text{ TONK } B, \quad A \text{ TONK } B / B,$

from which we obtain the schema A / B , licensing the inference of any sentence from any other, as a derived rule of inference! In order to avoid this sort of pathology, Belnap (1961) imposed the requirement that adding a logical constant to a given deductive system S by reference to a set of rules of inference result in a ‘conservative’ extension of S ; one in which no formula in the notation of S is provable that is not already provable in S .

Even thus amended, however, the syntactic conception of what it is to be a logical constant would appear to be too liberal. Let S be a standard system of proof for elementary logic, and suppose that P represents a fixed nonlogical predicate. Consider the quantifier Q such that $(Qx)(Ax)$ holds in any interpretation J if and only if some object falling under P satisfies $A(x)$ in J . (Thus, for example, if P represents ‘...is a bird’, the quantificational expression Q means ‘For some bird...’.) Let S^+ result from S by adding the following introduction and elimination rules for Q :

$$A(t), P(t)/(Qx)A(x) \quad (Qx)A(x)/A(c), P(c),$$

where in the second rule c is an individual constant subject to the non-occurrence requirement. It may be shown that S^+ is a conservative extension of S , but the expression Q thus characterized is not a logical constant.

Belnap's requirement may be seen as a by-product of a later viewpoint on syntactic characterizations of logical notions developed by Ian Hacking (1979). On this view, the fundamental rules of inference associated with the logical constants implicitly define their semantic interpretation. The three rules given above for conjunction determine its semantic interpretation in the following sense: there is precisely one truth table for 'Čamp;' with respect to which these rules are sound. The introduction and elimination rules for the existential quantifier can similarly be shown to characterize its semantic interpretation. It is clear that the indicated rules for Q do not thus characterize its semantic role, for its interpretation over any set varies with the interpretation assigned to P . Of course, to speak of a set of rules of inference as 'characterizing' the interpretation of a connective or a quantifier symbol presupposes that such an expression is to have an interpretation of the indicated (classical) sort; the justification of this assumption must be given in other terms. The idea is that in ascribing such an interpretation to an expression we look for the one - or a one - that makes the introduction and elimination rules associated with the expression truth-preserving; the expression counts as logical if over any universe there is only one such interpretation.

What is the relation between the semantic and syntactic solutions to the demarcation problem? One answer is that they correspond to two broad conceptions of the nature of logic (see [Tharp 1975](#)). On the first conception, logic is a theory of a certain type of deductive inference. On this view, what is distinctive about logical notions is their role in deductive inference, and logical consequence consists in a (perhaps idealized) type of deducibility. The syntactic view (on which, under very general conditions, a completeness result links a relation of logical consequence to the deductive system that characterizes the logical constants) provides a natural interpretation of this conception. An alternative conception, of more modern vintage, sees logic as an instrument for the characterization of structure. A logic (that is, a choice of logical constants) may be said to 'characterize' a structure if there is a set S of sentences in the logic such that any model of S exemplifies that structure. On this conception, the invariance criterion would seem to be very natural, for the behaviour of logical notions should be preserved under 'structure-preserving' transformations of any subject matter to which they are applied. From this point of view, there will be no particular premium on completeness, and the Löwenheim-Skolem property is an anomaly. Elementary logic is a paradigm of the first conception of logic and second-order logic typifies the second.

See also: [Logical and mathematical terms, glossary of](#); [Logical form](#); [Logical laws](#); [Proof theory](#)

TIMOTHY MCCARTHY

References and further reading

- Barwise, J. and Feferman, S.** (1985) *Model-Theoretic Logics*, Berlin: Springer. (An excellent sourcebook in generalized model theory, containing detailed treatments of all the technical issues touched on in this entry.)
- Belnap, N.** (1961) 'Tonk, Plonk, and Plink', *Analysis* 22: 130-4. (Addresses problems with syntactic criteria for interpreting logical constants in terms of introduction and elimination rules.)
- Etchemendy, J.** (1990) *The Concept of Logical Consequence*, Cambridge, MA: Harvard University Press. (An important critical evaluation of the Tarskian paradigm, emphasizing problems pertaining to the necessitation requirement.)
- Hacking, I.** (1979) 'What is Logic?', *Journal of Philosophy* 76 (6): 499-523. (Develops a sophisticated proof-theoretic account of the nature of logical constants.)
- Harman, G.** (1986) 'The Meaning of Logical Constants', in E. Lepore (ed.) *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, Oxford and New York: Blackwell, 125-34. (Explains the meaning of logical constants in terms of 'immediate implications' and 'immediate incompatibilities'.)
- Keisler, H.J.** (1971) 'Logic with the Quantifier "there exist uncountably many"', *Annals of Mathematical Logic* 1: 1-93. (Gives a completeness result for logic with the 'uncountability' quantifier, by means of a particularly nice axiomatization.)
- Klein, F.** (1872) *Vergleichende Betrachtungen über neuere geometrische Forschungen*, Erlangen: Deichert; repr. in *Gesammelte mathematische Adhandlungen*, Berlin: Springer, 1921, vol. 1, 460-97; trans. M.W. Haskell, 'A Comparative Review of Recent Geometric Researches', *Bulletin of the New York Mathematical Society* 2:

- 215-49, 1993. (The inspiration for Tarski 1966; of considerable interest in itself.)
- Lindström, P.** (1969) 'On Extensions of Elementary Logic', *Theoria* 31: 1-11. (A seminal article in generalized model theory, establishing that elementary logic is characterized by the completeness, compactness and Löwenheim-Skolem properties.)
- Löwenheim, L.** (1915) 'Über Möglichkeiten im Relativkalkül', *Mathematische Annalen* 76: 447-70; trans. S. Bauer-Mengelberg, 'On Possibilities in the Calculus of Relatives', in J. van Heijenoort, *From Frege to Gödel: A Sourcebook in Mathematical Logic*, Cambridge, MA: Harvard University Press, 1967. (A seminal paper in the theory of models, giving the first statement and proof of the Löwenheim-Skolem theorem for elementary logic.)
- McCarthy, T.** (1981) 'The Idea of a Logical Constant', *Journal of Philosophy* 78 (9): 499-523. (Discusses the invariance criterion and its problems, and includes an extension of the criterion to intensional logics.)
- McCarthy, T.** (1987) 'Modality, Invariance, and Logical Truth', *Journal of Philosophical Logic* 16: 423-43. (Establishes a general result saying that on weak assumptions if a set of notions is invariant across a class of possible worlds, then a logic based on those notions will satisfy a necessitation requirement with respect to that class.)
- Peacocke, C.** (1987) 'Understanding Logical Constants', *Proceedings of the British Academy* 83: 153-200. (Explains what is special about logical constants in terms of the way they are understood.)
- Prior, A.** (1960) 'The Runabout Inference-Ticket', *Analysis* 21: 38-9. (Criticizes the thesis that logical constants are generally characterized in terms of introduction and elimination rules.)
- Sher, G.** (1989) 'A Tarskian Conception of Logic', *Pacific Philosophical Quarterly* 70: 341-68. (An illuminating discussion of Tarski's later views on the demarcation problem.)
- Sher, G.** (1991) *The Bounds of Logic*, Cambridge, MA: MIT Press/Bradford. (A general discussion of problems in the philosophy of classical logic from a model-theoretic point of view.)
- Tarski, A.** (1936) 'Über den Begriff der logischen Folgerung', *Actes du Congres International de Philosophie Scientifique* 7: 1-11; repr. 'On the Concept of Logical Consequence', in *Logic, Semantics, Metamathematics*, trans. and ed. J.H. Woodger, Oxford: Oxford University Press, 2nd edn, 1983. (An early important informal discussion of the semantic conception of logical consequence.)
- Tarski, A.** (1966) 'What are Logical Notions?', ed. and with intro. by J. Corcoran, *History and Philosophy of Logic* 7: 143-54. (An illuminating discussion of notions of invariance in relation to logic.)
- Tharp, L.** (1975) 'Which Logic is the Right Logic?', *Synthese* 31: 1-21. (An excellent introduction to the problems treated in this entry.)

Logical form

Consider the following argument: *All men are mortal; Socrates is a man; therefore, Socrates is mortal.* Intuitively, what makes this a valid argument has nothing to do with Socrates, men or mortality. Rather, each sentence in the argument exhibits a certain ‘logical form’ and those forms, taken together, constitute a pattern that guarantees the truth of the conclusion given the truth of the premises. More generally, the logical form of a sentence of natural language is what determines both its logical properties and its logical relations to other sentences.

The logical form of a sentence of natural language is typically represented in a theory of logical form by a well-formed formula in a ‘logically pure’ language whose only meaningful symbols are expressions with fixed, distinctly logical meanings (for example, quantifiers). Thus, the logical forms of the sentences in the above argument would be represented in a theory based on pure predicate logic by the formulas ‘ $\forall x(Fx \rightarrow Gx)$ ’, ‘ Fy ’ and ‘ Gy ’, respectively, where ‘ F ’ and ‘ G ’ are free predicate variables and ‘ y ’ a free individual variable. The argument’s intuitive validity is then explained by the fact that the logical forms of the premises formally entail the logical form of the conclusion. The primary goal of a theory of logical form is to explain as broad a range of such intuitive logical phenomena as possible in terms of the logical forms that it assigns to sentences of natural language.

1 Logical form in Aristotelian logic

Broadly speaking, the ‘logical form’ of a sentence of natural language is what determines both its logical properties and its logical relations to other sentences. Most notably, the logical form of a sentence S determines whether or not it is logically true, and the logical forms of the sentences in a set K together with the logical form of S jointly determine whether or not the latter is a logical consequence of the former.

Although particularly prominent in the twentieth century, due especially to the influence of Bertrand Russell, the idea of logical form can be traced back to Aristotle. In the so-called ‘traditional logic’ that stems from Aristotle’s work in the *Prior Analytics*, the central form of argument is the ‘syllogism’, that is, an argument consisting of two premises and a conclusion. For example:

(1) Every whale is a mammal.
Some carnivore is a whale.
 $\therefore$ Some carnivore is a mammal.

Aristotle was the first to recognize explicitly that the validity of a syllogism (that is, its conclusion being a logical consequence of its premises) is entirely independent of the common noun phrases, or ‘terms’, in its constituent sentences. Rather, each sentence in a valid syllogism such as (1) above exhibits a certain form and, taken together, those forms constitute a pattern that, of itself, guarantees the truth of the conclusion given the truth of the premises.

To capture this idea systematically, Aristotle first restricted his attention to syllogisms the constituent sentences of which exhibit any of four basic sentential forms, known traditionally as *A*, *E*, *I* and *O*, respectively: ‘Every α is a β ’, ‘No α is a β ’, ‘Some α is a β ’ and ‘Some α is not a β ’, where ‘ α ’ and ‘ β ’ represent the roles of the subject and predicate terms in such sentences. Thus, the first premise of (1) is of the sentential form *A*, and both the second premise and the conclusion are of the form *I*. These sentential forms alone, however, are not enough to characterize the pattern that (1) exhibits. A further logically relevant feature of (1) is the way in which the terms occur in its constituent sentences: ‘carnivore’, for instance, is the subject term of both the second premise and the conclusion but does not occur at all in the first premise. To capture this feature generally, Aristotle introduced schematic variables F , G , H , ... Call the result of replacing the terms in a sentence that exhibits one of the basic sentential forms with distinct schematic variables a ‘schematic’ form. The pattern exhibited by a given syllogism can then be represented by replacing its constituent sentences with schematic forms that preserve the arrangement of terms in the syllogism. For example, (1) exhibits the following pattern:

(2) Every F is a G .
Some H is an F .
Some H is a G .

By sole means of representations such as (2), Aristotle proved, for each possible syllogistic pattern, whether or not it is valid; that is, for Aristotle, whether or not it is impossible for the premises of any instance of the pattern to be true and the conclusion false. Thus, the schematic forms of its premises and conclusion (relative to a given choice of schematic variables) completely determine the validity or invalidity of each pattern. Aristotle's schematic forms are thus paradigms of logical forms.

The central goal of a theory of logical form is to explain the logical properties of (and relations between) the members of a broad class *K* of sentences of natural language in terms of the logical forms that the theory assigns to those sentences. With only four general types of logical form to choose from - the four basic sentential forms, the scope of Aristotle's account as it stands is far too limited to count as a fully fledged theory of logical form. However, a perusal of any modern text with a section on traditional logic yields a variety of techniques for translating sentences with entirely different grammatical forms into instances of Aristotle's four basic sentential forms. For example, a simple individual assertion such as

(3) Matthew is a politician

is obviously not an instance of any of the basic sentential forms, since it does not begin with one of the quantifiers 'Every', 'Some' and 'No'. But by introducing a term that picks out a class containing only Matthew, (3) can be translated into the intuitively equivalent, albeit somewhat stilted, sentence

(4) Every person identical to Matthew is a politician,

which exhibits the sentential form A. Again, a sentence such as

(5) Someone gave *The Brothers Karamazov* to Andrea,

which involves a transitive verb, can be put into the appropriate sentential form by replacing the verb with its nominal '-er' counterpart followed by the preposition 'of', to produce a term that picks out exactly the class of individuals to which the original predicate applies. (5) then gives way to

(6) Some person is a giver of *The Brothers Karamazov* to Andrea.

Supplemented by such techniques, Aristotle's logic has a considerably broader application, as a much larger class of sentences can be assigned logical forms than would initially appear to be the case. So supplemented, then, Aristotle's account is plausibly taken to be the first genuine theory of logical form (see [Logic, ancient](#) §§1-3).

2 A general characterization of logical form

Although differing considerably in content and detail, the basic components and characteristics of Aristotle's theory of logical form (in this somewhat anachronistic rendering) are essentially the same as in contemporary accounts deriving from Bertrand [Russell](#). In this section, these components and characteristics are made explicit.

In general, the notion of logical form is always relative to a theory *T* of logical form which is directed towards some broad class *K* of sentences of natural language. Call *K* the 'target class' of the theory. The logical forms of the members of *K* are represented by formulas in a 'logically pure' canonical language; that is, a language that contains (perhaps in addition to punctuation) only variables and logical constants - expressions with fixed and, according to the theory, distinctly logical meanings (see [Logical constants](#) §1). For Aristotle, this is the language of schematic forms the logical constants of which are 'Every', 'Some', 'No', 'is a' and 'is not a'; and for Russell and his followers it is generally (some variant of) the language of *Principia Mathematica* (1910-13), with its now familiar quantifiers and propositional connectives.

The route from the sentences of the target class *K* to their logical forms according to *T* consists of two steps. First, and most important, is the translation of the members of *K* into an 'impure' hybrid language that consists of the canonical language supplemented by nonlogical constants: for Aristotle, ordinary terms such as 'carnivore' and also stilted terms such as 'person identical with Matthew'; for Russellians, individual constants such as 'Socrates' and *n*-place predicates such as 'mortal' (or perhaps abbreviations thereof such as 's' and 'M'). These nonlogical constants carry the same meaning as their informal counterparts and are used to construct the hybrid language translations of the sentences in which those counterparts occur. The translation of a sentence of *K* into a hybrid

language is typically known as its ‘analysis’. (Sentences such as ‘Every boy danced with a girl’, which are logically ambiguous, should of course receive a distinct analysis for each possible reading.) The analysis of a sentence is a paraphrase of the sentence which displays its logical form overtly in its surface grammatical structure. Thus, (3) and (5) are translated as (4) and (6), respectively, while for the Russellian, (7), for example, is translated as (8):

(7) If every man is mortal and Socrates is a man, then Socrates is mortal.

(8) $(\forall y(\text{man}(y) \rightarrow \text{mortal}(y)) \ \& \ \text{man}(\text{Socrates})) \rightarrow \text{mortal}(\text{Socrates})$.

In the second step of the process, the logical forms for the sentences of *K* are derived simply by uniformly replacing all the nonlogical constants in the analyses of those sentences with variables of the appropriate type. So, for example (assuming a choice of replacement variables for the nonlogical constants), the sentences in (1) - which in this case are identical with their analyses - yield the logical forms in (2), and (8) likewise yields (9) as the logical form of (7):

(9) $(\forall y(Fy \rightarrow Gy) \ \& \ Fx) \rightarrow Gx$.

As noted, the primary goal of a theory *T* of logical form is to account for the logical properties of the sentences in its target class *K* in terms of their logical forms. This is accomplished by means of a logical theory for the canonical language of *T* in which formal correlates of the ordinary logical notions - logical truth, logical consequence and so on - are defined for the formulas of the language. The apparent logical properties of the members of *K* are then explained (or explained away) by virtue of the fact that the formal correlates of those properties hold (or fail to hold) among the logical forms of the members of *K*. Thus, once again, the intuitive validity of (1) is explained by the formal validity of (2). Similarly, for the Russellian, the intuitive logical truth of (7) is explained by the formal logical truth of (9) - typically understood as truth in any domain under any interpretation (relative to that domain) of the free variables ‘*F*’, ‘*G*’ and ‘*x*’, following the work of Tarski (1933; see [Model theory](#)). In either case, the actual meanings of the nonlogical expressions in the natural language sentences in question play no role in determining the logical properties of those sentences. Rather, as the theories demonstrate, the logical properties are determined by their logical forms alone.

3 Comparing theories of logical form

The central principle for comparing theories of logical form is the following:

CP *T'* is preferable (*prima facie*) to *T* if *T'* explains a greater range of logical behaviour than *T* without appeal to extralogical principles (or, ‘meaning postulates’).

To illustrate, intuitively (10) follows directly from (5):

(10) Someone gave something to Andrea.

And yet, in Aristotle’s theory, as noted, (5) must be analysed along the lines of (6) and (10) along the lines of (11):

(11) Some person is a giver of something to Andrea.

However, then the analyses (6) and (11) exhibit only the obviously invalid argument pattern

(12) Some *A* is a *B*.

Some *A* is a *C*.

To capture the original entailment, (12) must be treated as enthymematic: that is, there is a suppressed premise with the logical form ‘Every *B* is a *C*’, namely

(13) Every giver of *The Brothers Karamazov* to Andrea is a giver of something to Andrea.

Aristotle’s logic, therefore, explains the relation between (5) and (10) by introducing an extralogical principle linking the two predicate terms, in effect reducing it to an instance of the argument pattern (2).

By contrast, represented in predicate logic, the logical forms of (5) and (10) constitute the valid pattern

(14) $\exists x Hxyz$
 $\exists y \exists x Hxyz$

Hence, that (10) is a consequence of (5) is explained on the basis of their logical forms alone, without appeal to any extralogical principles. So, *prima facie*, a theory based on predicate logic is preferable to Aristotle's theory.

The '*prima facie*' qualification in CP marks the influence of other factors besides extent of explanatory range that are relevant to an evaluation of competing theories of logical form. Two particularly influential principles can be observed in recent debates over the logical forms of one or another class of sentences: logical conservatism and ontological conservatism.

The shift from Aristotle to Russell exhibits a typical way in which one theory T' of logical form can broaden the explanatory range of another T ; namely, by extending or otherwise subsuming the canonical language and logic of T . In such cases, the language of T' is able to provide logical forms for all sentences in the target class of T . In addition, the language of T' includes logical constants and classes of complex expressions not present in the language of T that enable T' to construct logical forms not available to T . Appropriate logical principles for its new forms then enable T' to explain logical phenomena that are left unexplained (without extralogical principles) in T , for example, (10)'s being a logical consequence of (5). The principle of logical conservatism stipulates that supplementation to a given canonical language should be kept to a minimum. More exactly, let T and T' be theories with roughly equal target classes and let L be a background logical theory (first-order logic, for example) that is common to both T and T' ; then

LC T' is preferable (*prima facie*) to T if T' requires fewer extensions to L than T .

The increase in explanatory range of predicate logic, of course, is far too vast for LC to override CP in the choice between Aristotle and Russell. A better example is found in recent well-known work on the logical form of action sentences with adverbial modifiers. For example:

(15) April kissed Jonathan tenderly.

It is intuitively clear that (15) entails

(16) April kissed Jonathan.

Typical predicate logic analyses of these sentences represent 'kissed' and 'kissed tenderly' as distinct two-place predicates. The entailment is then explained by means of a meaning postulate to the effect that if A kisses B tenderly, then A kisses B . As with Aristotle's explanation of the relation between (5) and (10), then, in this explanation the apparent entailment is actually enthymematic. However, the entailment can be explained directly by treating an adverb such as 'tenderly' as an adverbial operator t that, when prefixed to an n -place predicate F (such as 'kissed'), yields a new n -place predicate $[tF]$ ('tenderly kissed') (see [Adverbs §1](#)). The logic of these new constructions is then characterized generally by the principle that

(17) $[\alpha\Pi]x_1 \dots x_n \rightarrow \Pi x_1 \dots x_n$, for any adverbial operator α and n -place predicate Π .

The logical forms of (15) and (16), then, on this approach, are (18) and (19), respectively:

(18) $[tF]xy$
 (19) Fxy

and so by the new logical principle (17), (16) follows from (15) directly in virtue of their logical forms.

Most philosophers would probably agree that this increase in explanatory range is significant enough to override LC and warrant the added apparatus. However, Davidson (1967) proposed a logically more conservative analysis which avoids the new apparatus. Specifically, for Davidson, the proper analysis of action sentences takes the structure of an action verb such as 'kissed' to involve an implicit parameter for an event. Thus, (15) is to be analysed as

(20) $\exists x$ (kissing-of(April, Jonathan, x) & tender(x))

(read, roughly, as ‘There is an event x such that x is a kissing of Jonathan by April and x is tender’), and (16) as

(21) $\exists x$ (kissing-of(April, Jonathan, x))

The entailment from (15) to (16) is then explained by standard logical principles governing conjunction and the existential quantifier, and the apparently hasty introduction of new constructions and unfamiliar logical principles is avoided. Since their explanatory ranges are the same, then, CP provides no support for the adverbial operator account over Davidson’s, and so, by LC, Davidson’s account is to be preferred (see [Adverbs](#)).

As Davidson’s account illustrates, however, conservatism with regard to logic is often accompanied by liberalism with regard to ontology - in this case, the postulation of events. The ‘ontological commitments’ of a theory T of logical form consist of the kinds of things that must exist if the analyses that T assigns to the sentences of its target class are to be meaningful (see [Ontological commitment](#)). The second principle - ontological conservatism - is that such commitments are to be kept to a minimum; that is, more exactly, where again T and T' have roughly equal target classes,

OC T' is preferable (*prima facie*) to T if T' has fewer ontological commitments than T .

Unlike LC, this principle favours the operator account of adverbial modification over Davidson’s more ontologically permissive account. The choice between the two thus turns on one’s preferred brand of conservatism.

LC tends to be overridden by CP when they conflict, as additional apparatus is usually viewed as a small price to pay for greater explanatory range. However, more is often at stake in conflicts between CP and OC, as theories of logical form that have great explanatory range often exact a high price in ontological commitment: possibilities (that is, possible but non-actual objects) are introduced to explain sentences involving modality (see [Modal logic](#), [philosophical issues in §§1-2](#)), intensional entities to explain the logic of attitude verbs (see [Propositional attitude statements §§1-2](#)) and so on. One must either choose between the two or offer a competing account with the same explanatory range but fewer ontological commitments. Since Russell first formulated his theory of descriptions (see [Descriptions](#)) to counter what he saw as Meinong’s ontological excesses, the development of competing theories of logical form has been, and largely remains, the primary tool for metaphysical discovery and the central approach to metaphysical debate in the twentieth century (see [Analytical philosophy](#)).

See also: [Logical and mathematical terms, glossary of](#); [Montague, R.M.](#); [Quine, W.V.](#)

CHRISTOPHER MENZEL

References and further reading

- Aristotle** (c.350s-330s BC) *Prior Analytics*, trans. and with commentary by R. Smith, Indianapolis, IN: Hackett Publishing Company, 1989, esp. I.1-7. (Referred to in §1. The primary source for Aristotle’s formal logic.)
- Davidson, D.** (1967) ‘The Logical Form of Action Sentences’, in N. Rescher (ed.) *The Logic of Decision and Action*, Pittsburgh, PA: University of Pittsburgh Press, 81-95; repr. with replies to critics in D. Davidson, *Essays on Actions and Events*, Oxford: Clarendon Press, 1980, esp. 137-46. (The source of Davidson’s account of action sentences with adverbial modifiers; discussed in §3.)
- Harman, G.** (1972) ‘Logical Form’, *Foundations of Language* 9: 38-65; repr. in D. Davidson and G. Harman (eds) *The Logic of Grammar*, Encino, CA: Dickenson, 1975, 289-307. (A fine discussion of theories of logical form on which the comparative principles in §3 are based.)
- Russell, B.A.W.** (1914) ‘Logic as the Essence of Philosophy’, in *Our Knowledge of the External World*, London: Allen & Unwin, 33-53; London: Routledge, 1993. (Perhaps Russell’s strongest statement of his view of the role of logical form in logic and philosophy.)
- Russell, B.A.W.** (1919) *An Introduction to Mathematical Philosophy*, London: Routledge, 1993. (An engaging if not always wholly clear account of the basic issues in philosophical logic and the philosophy of mathematics. Chapter 16 is a classic account of Russell’s theory of definite descriptions; chapter 18 includes a lengthy discussion of logical form.)
- Tarski, A.** (1933) *Pojęcie prawdy w językach nauk dedukcyjnych*, Warsaw; trans. J.H. Woodger (1956), ‘On the Concept of Truth in Formalized Languages’, in *Logic, Semantics, Metamathematics*, ed. J. Corcoran, Indianapolis, IN: Hackett Publishing Company, 2nd edn, 1983, 152-278. (Referred to in §2. The classic work of

model-theoretic semantics.)

Whitehead, A.N. and Russell, B.A.W. (1910-13) *Principia Mathematica*, Cambridge: Cambridge University Press, 3 vols; 2nd edn, 1925-7.(Referred to in §2. Probably the most influential work in logic in the twentieth century; extremely technical.)

Wittgenstein, L.J.J. (1922) *Tractatus Logico-Philosophicus*, trans. C.K. Ogden and F.P. Ramsey, London: Routledge; trans. D.F. Pears and B.F. McGuinness, London: Routledge, 1975.(A unique, highly influential work by Russell's most famous student in which a very interesting, somewhat eccentric notion of logical form plays a prominent role.)

Logical laws

There are at least three different kinds of answer to the question ‘What is a logical law?’ One establishes what it means for something to be a logical law. This answers the semantic question: What is the meaning of ‘logical law’? The second explains what makes something a logical law. This answers the metaphysical question: What is the ground of logical law? The third tells you what the logical laws are. This answers the question: What is the extension of ‘logical law’?

Even though logic is often seen as a complete science, the answers to all three questions are disputed. For example, there are at least three different conceptions of what it means for something to be a law of logic. Different conceptions account for logic in terms of necessity, truth in all models, and proof.

There are also different answers to the metaphysical question. If truth-preservation is central to logic, then the ground of logic depends on the metaphysics of truth. If logic is a matter of the meanings of terms, then the metaphysics of meaning is important for logic. Unfortunately, there is no widespread agreement on the metaphysics of meaning or truth.

Finally, there is no widespread agreement as to what the logical laws are. There are two general disputes here. First, it is not clear what notions count as logical. Does logic contain laws about identity, second-order quantification or modality? Second, given agreement on the scope of logic, there are still questions about the logical laws in that area. Intuitionists, quantum logicians, relevance and paraconsistent logicians each reject things taken as laws by others, even in the language of ‘and’, ‘or’ and ‘not’.

1 The semantic question

When we claim that something is a logical law, what do we mean? Sometimes we use ‘logical law’ to mean ‘theorem of a formal system’. Depending on taste, we might call theorems of first-order classical logic, or some other formal system, laws of logic. This is a derivative use of the phrase, akin to calling the claims of a particular scientific theory laws of nature. The propositions of a theory may be laws of nature, but only if the theory ‘gets it right’. The theory must describe the world in the right way for its claims to be laws of nature. Similarly, theorems of a formal system are only logical laws if they ‘get it right’ in some appropriate way. The interest consists in elaborating what it means for a theory to ‘get it right’. So, by ‘logical law’ we do not simply mean ‘theorem in a particular formal system’.

Another non-answer to this question is that logical laws are the ways we cannot help but think. Anyone familiar with the history of logic in the twentieth century will be aware that almost any principle that some take to be a logical law, others reject as invalid. If a logical law is something to which no one can help but assent, nothing counts as a logical law. No purely psychological answer, in terms of ‘laws of thought’, will give us logical laws.

Now to a more plausible account. The goal in the study of logic is an account of deductively valid inferences: these are the logical laws. A deductively valid inference is one in which necessarily, if the premises are true, so is the conclusion. Laws of logic are those inferences that are necessarily truth-preserving. This conception goes back to Aristotle (see [Logic, ancient §§1-3](#)).

A syllogism is a form of words in which when certain assumptions are made, something other than what has been assumed necessarily follows from the fact that the assumptions are such.

(Prior Analytics: 24b18)

We are interested in necessarily truth-preserving inferences because given these, and given true premises from which to infer, we will never (and in fact, *can* never) deduce falsehood. Given our concern for finding truth and avoiding falsehood, it is easy to see why logical laws are interesting, on this picture. It is also clear how logic can be normative. Given that we have the goal of gaining truth and avoiding falsehood, we *ought* to deduce validly. (Which is not to say that we ought not reason in other ways as well.)

We have one answer to our semantic question: a logical law is a necessarily truth-preserving inference. However, many have found this answer unsatisfactory. The main reason for dissatisfaction with this account is the reliance upon the intensional notion of necessity. Some have sought to give an account of logical laws that has no recourse

to necessity, or to any other intensional notion.

There are two important analyses of logical consequence which, at least at first glance, avoid using the notion of necessity. One was given its canonical exposition by Tarski in his 1936 essay ‘Über den Begriff der logischen Folgerung’ (‘On the Concept of Logical Consequence’), though it was prefigured by Bolzano in 1837. Tarski wrote ‘The sentence *X* follows logically from the sentences of the class *K* if and only if every model of the class *K* is also a model of the sentence *X* ([1936] 1956: 417; original emphasis).

There are no modal notions here, provided that there is an adequate non-modal explanation of what it is to be a model. There are a number of ways to do this, and there is no need for us to consider them in detail. Suffice to say that models are structures in which each of the non-logical constants in a language is given an interpretation in terms of the objects in the model, and the logical constants have a fixed interpretation in each model (see Logical constants §1; Model theory).

Model-theoretic notions of consequence do not, as they stand, make any appeal to modality. However, they do not give a modal-free account of necessary truth-preservation.

For example, the model-theoretic account needs a collection of logical terms. If *all* terms are logical, then the interpretation of each term is fixed, and there is only one model. This would make logical truth vacuous, as Tarski realized (1936). Now suppose that the terms ‘blue’ and ‘coloured’ are not counted as logical terms. The deduction ‘My car is blue; therefore my car is coloured’ fails to be valid on the model-theoretic account, because we can assign disjoint extensions to the predicates ‘blue’ and ‘coloured’. In contrast, the inference is necessarily truth-preserving. In this way, necessary truth-preservation and model-theoretic validity come apart. (For a detailed account of how the notions differ, see Etchemendy’s *The Concept of Logical Consequence* (1990).)

These two approaches to analysing logical validity differ quite sharply from a third account. According to the proof-theoretic account of validity, an argument is valid just when there is a proof of the conclusion from the premises. This approach is championed by Prawitz (1974) with many other constructivists, and Wagner (1987) (a non-constructivist) who sees this view in the work of Gottlob Frege. The proof-theoretic account grounds validity in the meanings of the logical constants. These meanings determine the validity of simple deductions (such as conjunction elimination: from *p* & *q* to derive *q*). An argument is valid if and only if there are simple deductions from the premises to the conclusion (see Proof Theory).

On this account, logical validity is analytic, and the epistemic function of logic (as a calculus for ideal justification) is obvious. As with model theory, a proof-theoretic account of validity depends on the choice of logical particles, and the rules that govern them.

The proof-theoretic account differs from the other two approaches. If validity is ultimately proof-theoretic, then the class of logical validities is recursively enumerable. If we can list the basic rules, we can list the valid deductions (provided that proofs are finite). Similarly, validity is compact (if *K* follows from *X*, then *K* must follow from some finite subset of *X*). Neither of these properties is essential to the other conceptions of logical validity.

Without doubt, each of these conceptions has a place in the study of logical law. It is harder to discern how they are to be related. My tentative proposal is this. Validity is a matter of necessary truth-preservation. Model-theoretic validity is important, because models can represent possibilities (however ontological commitment to possibilities is to be analysed). Sometimes, models do not represent real possibilities (as when the extension of ‘blue’ is not a subset of the extension of ‘coloured’) so, model-theoretic validity can differ from actual validity. Similarly, proof-theoretic validity can coincide with actual validity. Although analytic truth and necessary truth may not coincide in all cases, simple deductions involving logical particles are instances of necessary truth-preserving inferences. Perhaps these simple deductions will capture all of the necessary truth-preserving inferences in the language. In this case, the proof procedure is complete. On the other hand, the validities in the language may not be recursively enumerable, as in the case of second-order logic. In these cases, no recursive proof-theoretic account will capture all of the validities.

2 The metaphysical question

Even if we have decided what we mean when we call something a logical law, the issue of the *ground* of logical

law remains. If validity is necessary truth-preservation, an account has to be given of necessity and of truth. Similarly for the model-theoretic approach. If models are to bear any resemblance to the world, truth-in-a-model ought to have the same structure as truth *simpliciter*. In these cases, doctrines of truth are relevant to the outcome of logical theory. For example, intuitionists have claimed that, at least in the domain of mathematics, truth comes by way of construction or verification. This claim has resulted in disagreement about logical laws (see [Intuitionistic logic and antirealism](#)). The case is similar with quantum logic. Given a particular reading of the correspondence theory of truth (in which propositions correspond to subspaces of the Hilbert spaces of quantum physics), quantum logic seems to follow naturally (see [Quantum logic](#)). On the other hand, the correspondence, coherence or pragmatist theories of truth do not seem to dictate a particular theory of logic. Metaphysical doctrines of truth are relevant to an account of logic, but they need not determine that account.

Some take it that logical validity is independent of any particular account of truth, because logical laws are purely a matter of convention. Clearly convention has a large part to play in language, but it is much more than this to say that logical laws are simply true by convention. For example, it is clear that it is a matter of convention that ‘snow is white or snow is not white’ is used to express a truth (and hence, a logical truth). It is also a matter of convention that ‘snow is white’ refers to snow. In this case it would be very strange to say that the whiteness of snow is *purely* a matter of convention. Similarly, ‘snow is white or snow is not white’ is not true solely by convention, but also by the way the world is (see [Necessary truth and convention](#)).

Convention in logic is more at home in the view of logic as proof theory. Here, validity is a matter of the meanings of logical particles. Terms get meanings by convention. However, this does not exempt us from the difficult matters of semantics. Logic depends on the meanings of individual particles such as ‘not’.

Is the meaning of ‘not’ a matter of its truth-functionality, and does this yield classical negation? Or is negation to be analysed inferentially, with intuitionists? The analysis of meaning undergirds the proof-theoretic account of validity. No account of logical validity exempts the practitioner from the difficult task of giving an account of the ground of logical law. This is essentially a matter of metaphysics and semantics.

3 The question of extension

Even given answers to the semantic and metaphysical questions, we are not at an end. We have yet to give an account of which inferences qualify as logical truths. Logicians have a ready supply of formal systems designed to do just this: there is a whole wealth of different formal systems designed to give an account of valid inference in different languages. These systems differ from each other in a number of ways. They can contain the truth-functional connectives, quantifiers for objects (both the standard existential and universal quantifiers, and perhaps more exotic quantifiers), quantifiers of higher order, the identity relation, modalities of various sorts, and so on. If validity is construed as necessary truth-preservation, there is no need to make a principled choice of logical constants, because any term may be relevant in determining validities. The only choice is a pragmatic one, of which laws are worth studying. The way is open for a logic of colour, a logic of perception sentences, a logic of action, or a logic of all three. These ‘logics’ will be partial accounts of validity, just as a theory of electromagnetism is a partial account of physical law.

On the other accounts of validity, a distinction must be made between logical and non-logical terms, in order for there to be a univocal sense of logical validity. This is not an easy task, because there are no principled criteria that something must satisfy in order for it to be a logical constant. People disagree over whether generalized quantifiers of objects (such as ‘uncountably many’) and second-order quantifiers ought to feature in logic (see [Quantifiers, generalized; Second-order logic, philosophical issues in](#)).

Second, formal systems may differ in terms of their results, even if they have the same language. For example, intuitionistic logic and classical logic disagree about double negation elimination (not-not- p ; therefore p) provided that you take intuitionistic logic and classical logic to have the same vocabulary. There is a subtlety here. It is obvious that two different formal systems can be reconciled by treating them as modelling different things. Then intuitionistic logic and classical logic can be seen as not disagreeing, because they have different negations. Quine said as much when he claimed that changing a logic amounted to changing the subject (1970).

However, this is not the whole story. Classical logic and intuitionistic logic can disagree as to the validity of real arguments. Given a particular argument in natural language, an intuitionist and a classical logician may disagree

about its validity. If there is a fact of the matter as to whether the argument is valid, then classical logic and intuitionistic logic can be seen to disagree about this fact. The change of logic does not involve a change of subject when the subject (in this case, natural language arguments) is fixed in advance. Given a particular instance of the natural language 'not', classical logic and intuitionistic logic are different accounts of the valid arguments involving this fragment of natural language. The natural language particle is *prior* to its formalizations, which are intended to capture the meaning of the particle. These formalizations can capture the meaning well, or not well, and hence these formalizations can disagree.

With intuitionistic logic, the locus of disagreement is the conception of truth or, alternatively, the meanings of the logical particles. Given a verificationist or constructivist view of truth, you have a reason to be an intuitionist with regard to the standard logical particles. However, if you disagree with constructivism, and you do not analyse the meanings of logical terms in constructivist terms, you can use intuitionistic logic as a logic of necessary provability preservation. Disagreement about intuitionism can only be resolved by agreeing on theories of truth and meaning. The case is similar with quantum logic, as we saw before.

Disagreement becomes more subtle when we consider paraconsistency (permitting contradictions in some way to hold) and relevance. According to most logics, the argument from p & not- p to q is valid, but according to paraconsistent logics (including all logics of relevance - see [Anderson and Belnap 1975](#); [Anderson, Belnap and Dunn 1992](#)), it is not. In this disagreement no doctrines of truth or meaning stand out as motivating the peculiar features of the logics. Instead, there seem to be two major reasons that proponents of these systems can have for analysing validity in this way.

The first is quite simple. If a contradiction were true, or even possibly true, it would be clear that the inference from p & not- p to q would be invalid. While many doubt the coherence of this supposition, enough work has been done in the field of paraconsistent logic to show that the approach is coherent (see [Paraconsistent logic](#); [Priest, Routley and Norman 1989](#)).

If we admit the need to be able to deduce in inconsistent situations (whether they are epistemic situations, possible situations or actual situations), a logic without the inference from p & not- p to q is necessary. Note that this does not involve any particular doctrine about the nature of truth (correspondence, pragmatic or coherence), but simply the view that it is possible that a contradiction be true. Resolving such disagreements is a matter of examining the arguments for and against the truth of contradictions.

The second motivation for the view that the argument from p & not- p to q is invalid is quite different. On this view you need not hold that it is possible that a contradiction be true. Instead, you maintain that for an argument to be valid, and for a conditional to be true, its antecedent must be relevant to its consequent. If this is the case, then there may be a reason to reject the inference (see [Relevance logic and entailment](#)). Again, this disagreement is not about a particular doctrine of the nature of truth. Instead, it is a disagreement about the relationship between relevance on the one hand, and validity and conditionality on the other. Opponents can trade intuitions about the validity or otherwise of individual arguments, but this kind of discussion is rarely fruitful. A saner approach might go as follows: if the relevantist can develop a coherent theory of validity, which models our valid argument at least as well as can be done otherwise, and which, in addition, gives an account of phenomena that cannot be modelled otherwise, then, clearly, the relevantist position is viable and valuable. If it cannot, it will not succeed as a theory of validity.

In all of these considerations, we have seen that logic is quite similar to other sciences. The practitioners have some idea of what the subject matter is (valid inferences) but there is debate about its exact nature. The task is to model the subject matter, and there is no wide consensus about how this is to be done. However, as in other sciences, this does not mean that the study is not worthwhile, nor that it will not enrich our knowledge of the way things are.

See also: [Logical and mathematical terms, glossary of](#); [Logical form](#)

GREG RESTALL

References and further reading

Anderson, A.R. and Belnap, N.D., Jr (1975) *Entailment: The Logic of Relevance and Necessity*, vol. 1,

- Princeton, NJ: Princeton University Press. (Together with **volume 2**, this gives the canonical exposition of much of the work on logics of relevance implication and entailment.)
- Anderson, A.R., Belnap, N.D., Jr and Dunn, J.M.** (1992) *Entailment: The Logic of Relevance and Necessity*, vol. 2, Princeton, NJ: Princeton University Press. (Both volumes together give the canonical exposition of much of the work on logics of relevance implication and entailment. Volume 2 also includes an interesting discussion of different approaches to relevance validity.)
- Aristotle** (c. mid 4th century BC) *Prior Analytics*, trans. Richard McKeon, in *The Basic Works of Aristotle*, New York: Random House, 1941. (Aristotle's work on the syllogism.)
- Boolos, G.** (1975) 'On Second Order Logic', *Journal of Philosophy* 72: 509-27. (Defence of second-order logic as logic.)
- Etchemendy, J.** (1990) *The Concept of Logical Consequence*, Cambridge, MA: Harvard University Press. (Critique of the model-theoretic account of validity.)
- Prawitz, D.** (1974) 'On the Idea of a General Proof Theory', *Synthese* 27: 63-77. (Exposition and defence of the proof-theoretic account of validity.)
- Priest, G., Routley, R. and Norman, J.** (1989) *Paraconsistent Logic: Essays on the Inconsistent*, Munich: Philosophia. (Essays on paraconsistent logic.)
- Quine, W.V.** (1970) *Philosophy of Logic*, Englewood Cliffs, NJ: Prentice Hall. (Exposition and defence of the model-theoretic account of validity, with a discussion of the scope of logic, and 'deviant' logics.)
- Shapiro, S.** (1991) *Foundations Without Foundationalism*, Oxford: Oxford University Press. (Defence and applications of second-order logic.)
- Sher, G.** (1991) *The Bounds of Logic*, Cambridge, MA: MIT Press. (Discussion of first-order quantification and the scope of logic, from the perspective of model theory.)
- Tarski, A.** (1936) 'Über den Begriff der logischen Folgerung', *Actes du Congrès International de Philosophie Scientifique* 7: 1-11; trans. J.H. Woodger, 'On the Concept of Logical Consequence', in *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*, Oxford: Clarendon Press, 1956. (The classic exposition of the model-theoretic account of validity.)
- Wagner, S.** (1987) 'The Rationalist Conception of Logic', *Notre Dame Journal of Formal Logic* 28: 3-35. (Exposition and defence of the proof-theoretic account of validity, and first-order logic.)

Logical positivism

Logical positivism (logical empiricism, neo-positivism) originated in Austria and Germany in the 1920s. Inspired by late nineteenth- and early twentieth-century revolutions in logic, mathematics and mathematical physics, it aimed to create a similarly revolutionary scientific philosophy purged of the endless controversies of traditional metaphysics. Its most important representatives were members of the Vienna Circle who gathered around Moritz Schlick at the University of Vienna (including Rudolf Carnap, Herbert Feigl, Kurt Gödel, Hans Hahn, Karl Menger, Otto Neurath and Friedrich Waismann) and those of the Society for Empirical Philosophy who gathered around Hans Reichenbach at the University of Berlin (including Walter Dubislav, Kurt Grelling and Carl Hempel). Although not officially members of either group, the Austrian philosophers Ludwig Wittgenstein and Karl Popper were, at least for a time, closely associated with logical positivism.

The logical positivist movement reached its apogee in Europe in the years 1928-34, but the rise of National Socialism in 1933 marked the effective end of this phase. Thereafter, however, many of its most important representatives emigrated to the USA. Here logical positivism found a receptive audience among such pragmatically, empirically and logically minded American philosophers as Charles Morris, Ernest Nagel and W.V. Quine. Thus transplanted to the English-speaking world of 'analytic' philosophy it exerted a tremendous influence - particularly in philosophy of science and the application of logical and mathematical techniques to philosophical problems more generally. This influence began to wane around 1960, with the rise of a pragmatic form of naturalism due to Quine and a historical-sociological approach to the philosophy of science due mainly to Thomas Kuhn. Both of these later trends, however, developed in explicit reaction to the philosophy of logical positivism and thereby attest to its enduring significance.

1 Historical background

Immanuel Kant, in the positivists' eyes, had made a lasting contribution to scientific philosophy - particularly in his rejection of the possibility of supersensible metaphysical knowledge and his reorientation of theoretical philosophy around the two questions 'how is pure mathematics possible?' and 'how is pure natural science possible?' In answering these questions Kant developed his famous defence of synthetic a priori knowledge - knowledge independent of sensible experience yet nonetheless substantively applicable to the empirical world. For Kant, the mathematical physics of Newton paradigmatically exemplified such synthetic a priori knowledge through its reliance on Euclidean geometry and fundamental laws of motion such as the law of inertia. Kant's theory of a priori faculties of the mind - the faculty of pure intuition or sensibility and the faculty of pure understanding - was then intended to explain the origin of synthetic a priori knowledge and thus make philosophically comprehensible the possibility of Newtonian mathematical physics.

After the intervening dominance of post-Kantian idealism, a number of German-speaking philosophers renewed the call for a scientific, epistemological and non-metaphysical form of philosophy. But these Neo-Kantian philosophers also had to face an important new challenge to the Kantian synthetic a priori: the nineteenth-century development of non-Euclidean geometry by Gauss, Bolyai, Lobachevskii, Riemann and Klein (see [Geometry, philosophical issues in §§1, 3](#)). Although some Neo-Kantians attempted to defend the uniqueness and apriority of Euclidean geometry nonetheless, others - especially those of the Marburg School such as Paul Natorp and Ernst Cassirer - aimed to generalize the synthetic a priori beyond its particular embodiment in classical Euclidean-Newtonian mathematical physics (see [Neo-Kantianism](#)). This latter tendency was similar in important respects to ideas the logical positivists were to elaborate.

But the most important nineteenth-century predecessors of logical positivism were Hermann von Helmholtz, Ernst Mach and Henri Poincaré. Through their efforts to comprehend the radical changes sweeping through nineteenth-century science, these three thinkers initiated a new style of scientific philosophy later taken up and systematized by the positivists. The changes in question included the rise of non-Euclidean geometry, the formulation of the conservation of energy and general thermodynamics, and the beginnings of scientific physiology and psychology. Helmholtz made fundamental contributions to all three areas. He based geometry on the postulate of 'free mobility' of rigid bodies, and, since all classical geometries of constant curvature - negative, positive or zero (Euclidean) - satisfy this postulate, he opposed the Kantian commitment to the apriority of

geometry: whether space is Euclidean or non-Euclidean is an empirical question about the actual behaviour of rigid bodies. In physiology, Helmholtz articulated a general principle of psycho-physical correlation whereby our sensations correspond to - but are in no way pictures or images of - processes in the external physical world. These processes consist, in the end, of microscopic atoms interacting via central forces, and, on this basis, Helmholtz developed his famous interpretation of the conservation of energy.

Mach and Poincaré can be seen as reacting, in diverse ways, to Helmholtz. Mach attacked especially atomism and the idea of a psychophysical correlation between two incommensurable realms, and he advanced a programme for the unity of science based on immediately perceptible 'elements' or 'sensations'. The task of science consists solely in seeking correlations among such elements (as in phenomenological thermodynamics), and all dualistic and atomistic tendencies are to be purged as metaphysical via historico-critical analysis. This Machian empiricism exerted a decisive influence on the logical positivists. Poincaré, on the other hand, influenced the positivists primarily through his philosophy of geometry. He agreed with Helmholtz's emphasis on the free mobility of rigid bodies but disagreed with Helmholtz's empiricism. According to Poincaré, the idea of a rigid body is an idealization that cannot be straightforwardly instantiated in the physical world. By freely choosing one of the three classical geometries as, so to speak, a definition of rigidity, we then first make it possible to carry out empirical investigations with real physical bodies. Physical geometry is thus neither synthetic a priori nor empirical: it is 'conventional' (see [Conventionalism](#)).

2 Relativistic physics

Albert Einstein's special (1905) and general (1916) theories of relativity entered this volatile intellectual situation as a revelation (see [Einstein, A.; General relativity, philosophical responses to; Relativity theory, philosophical significance of](#)). And the relativistic revolution in physics directly stimulated [Schlick](#), [Reichenbach](#) and [Carnap](#) to initiate a parallel revolution in scientific philosophy. All three thinkers agreed that relativity - especially through the general relativistic description of gravitation via a (four-dimensional) geometry of variable curvature - definitively refutes the Kantian idea that Euclidean geometry is synthetic a priori. Moreover, relativity arises from critical reflection on the empirical significance of spatiotemporal concepts in physics (in particular, the concept of simultaneity and the concept of motion) and thus demonstrates the fruitfulness of Mach's basic point of view. At the same time, however, through its use of sophisticated abstract mathematics, relativity also illustrates the limitations of Machian empiricism (according to which even mathematical concepts have an empirical origin). All three thinkers therefore attempted to formulate an intermediate position that would do justice to both Machian empiricism and the continued importance of a priori mathematical elements in physics. Poincaré's concept of convention came to play a central role.

Schlick, Reichenbach and Carnap first pursued rather different paths. Whereas Schlick emphasized from the outset that the Kantian synthetic a priori has no place at all in the new relativistic context, Reichenbach and Carnap initially attempted to salvage important aspects of Kantianism. Reichenbach began by distinguishing the idea of necessary and unrevisable truth from the idea of necessary presupposition of a given scientific conceptualization of nature. For Reichenbach, relativity refuted the former but embodied the latter. Kant was right that the necessary presuppositions of Newtonian physics included Euclidean geometry and the laws of motion. In moving to relativistic physics, however, these are replaced by fundamentally new presuppositions. We thus end up with a relativized version of the Kantian a priori (as constituting the presuppositions of a particular theory). Carnap, by contrast, began by distinguishing metrical from topological features of physical space. The latter are indeed synthetic a priori as Kant thought (they even depend on a kind of pure intuition), but the former - as general relativity has shown - essentially involve the behaviour of empirically given bodies. We thus end up with a weakening of the Kantian a priori (from metrical to topological features).

These early attempts to salvage aspects of the synthetic a priori did not survive, however. For Schlick's view that relativity is simply incompatible with Kant eventually won the day. Although the distinction between Poincaré's conventionalism and Helmholtzian empiricism was not entirely clear (and Reichenbach, in particular, preferred to associate his later viewpoint with Helmholtz rather than Poincaré), both Reichenbach and Carnap soon came to replace the Kantian notion of the a priori with Poincaré's concept of convention. Yet this form of conventionalism (unlike Poincaré's) was forged in the crucible of a revolutionary new physics and thus demonstrated the vitality and relevance of a new philosophy.

3 Logic and the foundations of mathematics

Whereas the positivists appealed to Poincaré's concept of convention (as realized, so they thought, in relativistic physics) to give a new answer to Kant's question concerning the possibility of pure natural science, they appealed to modern developments in logic and the foundations of mathematics to give a new answer to Kant's question concerning the possibility of pure mathematics. There were in fact two distinguishable sets of developments here. The formal point of view, typified by David Hilbert's logically rigorous axiomatization of geometry, freed geometry from any reference at all to intuitively spatial forms and instead portrayed its subject matter as consisting of any things whatever that satisfy the relevant axioms (see [Geometry, philosophical issues in §2; Hilbert's programme and formalism](#)). Geometry is rigorously and a priori true, not because it reflects the structure of an intuitively given space, but rather because it 'implicitly defines' its subject matter via purely logical - but otherwise entirely undetermined - formulas. Mathematical truth, on this view, is identified with logical consistency. The 'logicism' of Gottlob Frege and Bertrand [Russell](#), by contrast, aimed to construct particular mathematical disciplines (especially arithmetic) within an all-embracing system of logic. On this view mathematical disciplines (like arithmetic) indeed have a definite subject matter about which they express truths: namely, the subject matter of logic itself (propositions, classes, and so on). As thus purely logical, however, such pure mathematical disciplines express merely analytic truths and are not synthetic a priori (see [Logicism](#)).

Hilbert's formal point of view was pursued especially by Schlick, who in a sense made the notion of implicit definition, together with the associated distinction between undetermined form and determinate (given) content, the centrepiece of his philosophy. The logicist point of view, by contrast, was pursued especially by Carnap, who studied with Frege and then was decisively influenced by Russell. Indeed, Carnap was inspired by Russell's conception of 'logic as the essence of philosophy' to reconceive philosophy itself on the model of the logicist construction of arithmetic. He began, in *Der logische Aufbau der Welt* (1928), by developing a 'rational reconstruction' of empirical knowledge - an epistemology - within the logical framework of Russell and Whitehead's *Principia Mathematica* (1910-13). By defining or 'constituting' all concepts of empirical science within this logic from a basis of subjective 'elementary experiences', Carnap's reconstruction was to show, among other things, that the dichotomy between empirical truth and analytic/definitional truth is indeed exhaustive.

Yet the logic of *Principia Mathematica* was afflicted with serious technical difficulties: the need for special existential axioms such as the axioms of infinity and choice. Partly in response to such difficulties, Ludwig Wittgenstein asserted in his *Tractatus Logico-Philosophicus* (1922) that logic has no subject matter after all: the propositions of logic are entirely tautological or empty of content (see [Wittgenstein, L.J.J. §§3-7](#)). Carnap eagerly embraced this idea, but he also attempted to adapt it to the new, post-*Principia* technical situation - which involved the articulation of the 'intuitionist' or 'constructivist' point of view by L.E.J. Brouwer and the development of meta-mathematics by Hilbert and Kurt Gödel (see [Intuitionism §§2-3; Logic in the early 20th century §§6-9; Mathematics, foundations of](#)). In *Logische Syntax der Sprache* (1934) Carnap formulated his mature theory of formal languages and put forward his famous 'Principle of Tolerance' - according to which logic has no business at all looking for true or 'correct' principles. The task of logic is rather to investigate the structure of any and all formal languages - 'the boundless ocean of unlimited possibilities' - so as to map out and explore their infinitely diverse logical structures. Indeed, the construction and logical investigation of such formal languages became, for Carnap, the new task of philosophy. The concept of analyticity thereby took on an even more important role. For this concept characterizes logical as opposed to empirical investigation and thus now expresses the distinctive character of philosophy itself.

4 The Vienna Circle

Otto [Neurath](#), Hans Hahn, and the physicist Philipp Frank initiated a discussion group in Vienna, beginning in 1907, in which they considered a combination of Machian empiricism with Poincaré's new insights into the conventional character of physical geometry. Deeply impressed by Schlick's work on relativity theory, they arranged (apparently with Einstein's help) to bring Schlick to the University of Vienna in 1922 to take over the Chair in Philosophy of the Inductive Sciences previously held by Mach. What we now know as the Vienna Circle quickly took shape. Reichenbach, who had become acquainted with Carnap through their common interest in relativity, introduced him to Schlick in 1924. In 1925 Carnap lectured to the Circle in Vienna on his new 'constitutional theory of experience' and became assistant professor under Schlick in 1926. The Circle then

engaged in intensive discussions of Carnap's epistemology and Wittgenstein's *Tractatus*. Wittgenstein's view that all propositions are truth-functions of 'elementary propositions' was combined with Carnap's constitution of scientific concepts from a basis of 'elementary experiences' so as to create a new, logically rigorous form of empiricism according to which all meaningful - scientific - propositions are reducible to propositions about immediately given experience. And this was articulated as the 'official' philosophy of the Vienna Circle in the famous manifesto *Wissenschaftliche Weltauffassung* in 1929.

Neurath was the driving force in thus turning the Vienna Circle into a public philosophical movement. Trained in economics and the social sciences, Neurath was extremely active politically as a scientific neo-Marxist. In particular, he took the community of natural scientists as the model for a rationally organized human society, and, on this basis, he advocated a reorganization of both intellectual and social life from which all non-rational, 'metaphysical' elements would be definitively purged. In this sense, Neurath saw the philosophical work of the Vienna Circle as a reflection of the wider movement for a *neue Sachlichkeit* then current in Weimar culture - as typified, for example, by the Dessau Bauhaus. As in the wider culture, this movement stood in philosophy for a rejection of individualism in favour of the cooperative, piecemeal, and 'technological' approach to problems exemplified in the sciences, and it was therefore particularly hostile to what was perceived as a return to the metaphysical system-building of post-Kantian idealism by influential German philosophers such as Martin Heidegger. Carnap was especially sympathetic to Neurath's broader philosophical-political vision and clearly expresses this vision in the Preface to the *Aufbau*. Schlick, by contrast, preferred a more individualistic model of philosophy and resisted the idea of a 'movement'.

This divergence between a 'left wing' and a 'right wing' of the Circle emerged in the sphere of epistemology in a debate over 'protocol-sentences' in the years 1930-4. At issue was the status of the basic propositions or protocols in which the results of scientific observation are recorded. It had initially appeared, in Carnap's constitutional system of the *Aufbau*, that such propositions must express private, subjective sense-experience. For Neurath, however, this view was inconsistent with the publicity and intersubjectivity required by science. He therefore advocated a more naturalistic conception of protocols as sentences accepted by the scientific community as recording the results of observation at a given time (see *Naturalized philosophy of science*). These sentences must thus be expressible within the public and 'physicalistic' language of unified science and hence, like all other sentences, are in principle revisable. Schlick was deeply shocked by Neurath's view - which he took to represent an abandonment of empiricism in favour of the coherence theory of truth (see *Truth, coherence, theory of*). Carnap attempted, in typical fashion, to mediate the dispute: at issue was simply a choice between two different languages in which to formulate or rationally reconstruct the results of unified science. Although Neurath's thoroughly intersubjective 'physicalistic' language (where, as Karl Popper emphasized especially, every sentence is revisable) was clearly preferable on pragmatic grounds, Carnap held that this choice - like every other choice of formal language - is in the end conventional (see *Popper, K.R. §3*). Empiricism, in Carnap's hands, is itself framed by conventional and hence non-empirical choices (see *Vienna Circle*).

5 Emigration, influence, aftermath

The rise of the Nazi regime set off a wholesale migration of logical positivists to the English-speaking world. Carnap, who had become professor at Prague in 1931, moved in 1936 to the University of Chicago. Reichenbach, who had fled to Istanbul in 1933, moved in 1938 to the University of California at Los Angeles. (After Reichenbach's death in 1953 Carnap took over his position at UCLA, beginning in 1954.) Neurath, after leaving Vienna for The Hague in 1934, fled to England in 1940 - where he worked in Oxford until his death in 1945. Friedrich Waismann fled for England as well, where he lectured at Oxford from 1939. Philip Frank emigrated to the USA (also from Prague) in 1938 and settled at Harvard in 1939. Karl Menger took up a position at Notre Dame in 1937, and Kurt Gödel became a member of the Institute for Advanced Study at Princeton in 1940. Herbert Feigl went first to the University of Iowa in 1933 and then to the University of Minnesota in 1940, where he founded the influential Minnesota Center for the Philosophy of Science in 1953. Carl Hempel joined Carnap at the University of Chicago in 1939 and, after teaching at Queens College and Yale, settled at Princeton in 1955. (Schlick was murdered by a deranged student at the University of Vienna in 1936.)

The growth of philosophy of science in the USA was decisively shaped by the work of Carnap, Reichenbach and Hempel. Reichenbach influenced especially the development of philosophy of physics through his work on

geometry, relativity and the direction of time (see [Relativity theory, philosophical significance of §3](#)). Hempel published extraordinarily influential papers on the logical analysis of explanation and confirmation and thereby furthered the ideal of scientific philosophy first articulated by Carnap (see [Explanation](#); [Confirmation theory](#)). Carnap himself continued the construction of formal languages in which such concepts as testability, modality and probability could be rationally reconstructed or ‘explicated’ and thus contributed further to the same ideal. Indeed, Carnap’s explication of concepts through the construction of formal languages influenced the English-speaking world of analytic philosophy far beyond the borders of philosophy of science. Developments in formal semantics and philosophy of language, in particular, rested on Carnap’s initial work on modality (see [Semantics §2](#)).

The Carnapian ideal of explication is based on a sharp distinction between logical and empirical investigation, analytic and synthetic truth. In his *Logical Syntax of Language* (1934) Carnap had attempted a general explication of the concept of analyticity itself - a general formal method for distinguishing, within the context of any given formal language, the analytic from the synthetic sentences of that language. After accepting Alfred Tarski’s semantical conception of truth in 1935, however, Carnap abandoned the approach of *Logical Syntax* and frankly admitted that (although explications for various particular languages could still be constructed) he now had no generally applicable explication of the concept of analyticity. After studying with Carnap in the early 1930s, W.V. Quine then exploited this situation to attack the concept of analyticity as such and, on this basis, to attack the Carnapian ideal of logical explication as well (see [Quine, W.V. §§2, 4](#)). Philosophy, for Quine, is itself a kind of empirical science - a branch of human psychology or ‘naturalized epistemology’ (see [Naturalized epistemology](#)). Moreover, at the same time that Quine was articulating this new philosophical vision, Thomas Kuhn published *The Structure of Scientific Revolutions* (1962) in the *International Encyclopedia of Unified Science* edited by Carnap and Charles Morris. Whereas Carnap had relegated the (conventional) choice of scientific language to the limbo of pragmatics, Kuhn concentrated on those factors - especially social factors - which, in a scientific revolution, determine precisely this kind of choice. These ideas, in harmony with Quine’s more general naturalistic point of view, then led to historical and sociological approaches to the study of science and thus, in the end, to the decline of logical analyses of scientific language in the Carnapian style.

See also: [Analysis, philosophical issues in](#); [Analyticity](#); [Bridgman, P.W.](#); [Emotivism](#); [Empiricism](#); [Logical atomism](#); [Meaning and verification](#); [Operationalism](#); [Positivism in the social sciences](#); [Theories, scientific](#); [Unity of science](#)

MICHAEL FRIEDMAN

References and further reading

- Ayer, A.** (ed.) (1959) *Logical Positivism*, New York: Free Press. (Very useful short collection. Contains, in particular, some of the most important papers from the ‘protocol-sentence’ debate discussed in §4.)
- Carnap, R.** (1922) *Der Raum: Ein Beitrag zur Wissenschaftslehre*, Berlin: Reuther & Reichard. (Carnap’s early analysis of relativity theory. Discussed in §2.)
- Carnap, R.** (1928) *Der logische Aufbau der Welt*, Berlin: Weltkreis; trans. R. George, *The Logical Structure of the World*, Berkeley, CA: University of California Press, 1967. (Referred to in §§3, 4.)
- Carnap, R.** (1934) *Logische Syntax der Sprache*, Vienna: Springer; trans. A. Smeaton, *The Logical Syntax of Language*, London: Routledge & Kegan Paul, 1937. (Referred to in §§3, 5. English version contains important material not appearing in the original.)
- Carnap, R.** (1947) *Meaning and Necessity*, Chicago, IL: University of Chicago Press. (Carnap’s most influential post-war contribution to logic and philosophy of language. Second edition appearing in 1956 contains important supplementary material.)
- Carnap, R.** (1950) *Logical Foundations of Probability*, Chicago, IL: University of Chicago Press. (Carnap’s fundamental analysis of ‘logical probability’ and the concept of degree of confirmation.)
- Carnap, R.** (1963) ‘Intellectual Autobiography’, in P. Schilpp (ed.) *The Philosophy of Rudolf Carnap*, La Salle, IL: Open Court. (Contains detailed information about the historical development of the logical positivist movement.)
- Coffa, A.** (1991) *The Semantic Tradition from Kant to Carnap*, Cambridge: Cambridge University Press. (Most extensive scholarly treatment of the background to and evolution of logical positivism.)
- Creath, R.** (ed.) (1990) *Dear Carnap, Dear Van: The Quine-Carnap Correspondence and Related Work*, Berkeley, CA: University of California Press. (Excellent source for the Quine-Carnap debate on analyticity discussed in §5.)

- Feigl, H.** (1969) 'The Wiener Kreis in America', in D. Fleming and B. Bailyn (eds) *The Intellectual Migration*, Cambridge, MA: Harvard University Press. (Discussion of the positivists' migration to America by a participant.)
- Frank P.** (1949) *Modern Science and its Philosophy*, Cambridge, MA: Harvard University Press. (Classic discussion of the positivist movement by a participant. Contains, in particular, good discussions of the influence of Mach and Poincaré.)
- Haller, R.** (1993) *Neopositivism*, Darmstadt: Wissenschaftliche Buchgesellschaft. (Special attention to Neurath.)
- Hempel, C.** (1965) *Aspects of Scientific Explanation*, New York: Free Press. (Collection of Hempel's classic papers which decisively shaped philosophy of science in the USA.)
- Kraft, V.** (1950) *Der Wiener Kreis*, Vienna: Springer; trans. A. Pap, *The Vienna Circle*, New York: Philosophical Library, 1953. (Good discussion of how Carnap's *Aufbau* and Wittgenstein's *Tractatus* were combined by the Vienna Circle.)
- Kuhn, T.** (1962) *The Structure of Scientific Revolutions*, Chicago, IL: University of Chicago Press. (Referred to in §5. Extraordinarily influential reaction to logical positivist philosophy of science.)
- Neurath, O.** (1981) *Gesammelte philosophische und methodologische Schriften*, eds R. Haller and H. Rutte, Vienna: Holder-Pichler-Tempsky. (Most extensive collection of Neurath's philosophical works in the original. Contains the manifesto *Wissenschaftliche Weltauffassung* on pages 299-336.)
- Neurath, O.** (1983) *Philosophical Papers 1913-1946*, ed. and trans. R. Cohen and M. Neurath. (Most extensive collection of Neurath's philosophical works in English.)
- Popper, K.** (1935) *Logik der Forschung*, Berlin: Springer; *The Logic of Scientific Discovery*, London: Hutchinson, 1958. (Popper's extraordinarily influential discussion of scientific method - partly in the spirit of, but also partly explicitly opposed to, logical positivism.)
- Reichenbach, H.** (1920) *Relativitätstheorie und Erkenntnis Apriori*, Berlin: Springer; trans. M. Reichenbach, *The Theory of Relativity and A Priori Knowledge*, Berkeley, CA: University of California Press, 1965. (Reichenbach's earliest analysis of relativity theory. Discussed in §2.)
- Reichenbach, H.** (1928) *Philosophie der Raum-Zeit-Lehre*, Berlin: de Gruyter; trans. M. Reichenbach and J. Freund, *The Philosophy of Space and Time*, New York: Dover, 1958. (Reichenbach's most influential analysis of geometry and relativity theory.)
- Reichenbach, H.** (1951) *The Rise of Scientific Philosophy*, Berkeley, CA: University of California Press. (Widely read popularization.)
- Reichenbach, H.** (1978) *Selected Writings, 1909-1953*, ed. M. Reichenbach and R. Cohen, Dordrecht: Reidel. (English collection of Reichenbach's most important papers.)
- Russell, B. and Whitehead, A.** (1910-13) *Principia Mathematica*, Cambridge: Cambridge University Press. (Referred to in §3. Monumental presentation of modern mathematical logic. Exerted a profound influence on logical positivism.)
- Schlick, M.** (1917) *Raum und Zeit in der gegenwärtigen Physik*, Berlin: Springer; trans. H. Brose, *Space and Time in Contemporary Physics*, Oxford: Oxford University Press, 1920. (Schlick's widely read and influential treatment of relativity theory. Discussed in §2.)
- Schlick, M.** (1918) *Allgemeine Erkenntnislehre*, Berlin: Springer; trans. H. Feigl and A. Blumberg, *General Theory of Knowledge*, La Salle, IL: Open Court, 1974. (Schlick's major epistemological work. Contains, in particular, the general theory of 'implicit definition' discussed in §3. Second edition appearing in 1925 - on which the English translation is based - is significantly revised.)
- Schlick, M.** (1938) *Gesammelte Aufsätze*, Vienna: Gerold. (Collection of papers, 1926-36.)
- Schlick, M.** (1979) *Philosophical Papers*, eds H. Mulder and B. van de Velde-Schlick, Dordrecht: Reidel. (Most extensive collection of Schlick's works in English. Contains, in particular, the Brose translation of Schlick 1917.)
- Uebel, T.** (1992) *Overcoming Logical Positivism from Within*, Amsterdam: Rodopi. (Most extensive treatment - with special emphasis on Neurath's role - of the 'protocol-sentence' debate discussed in §4.)
- Waismann, F.** (1967) *Wittgenstein und der Wiener Kreis*, Frankfurt: Suhrkamp; trans. J. Schulte and B. McGuinness, *Wittgenstein and the Vienna Circle*, Oxford: Blackwell, 1979. (Very useful material on Wittgenstein's special influence.)
- Wittgenstein, L.** (1922) *Tractatus Logico-Philosophicus*, trans. C. Ogden with an introduction by B. Russell, London: Routledge & Kegan Paul. (Referred to in §§3, 4. Although Wittgenstein himself was not officially a member, this work and Carnap's *Aufbau* constitute the two most central texts of the logical positivist movement. New translation D. Pears and B. McGuinness, 1961.)

Logicism

The term 'logicism' refers to the doctrine that mathematics is a part of (deductive) logic. It is often said that Gottlob Frege and Bertrand Russell were the first proponents of such a view; this is inaccurate, in that Frege did not make such a claim for all of mathematics. On the other hand, Richard Dedekind deserves to be mentioned among those who first expressed the conviction that arithmetic is a branch of logic.

The logicist claim has two parts: that our knowledge of mathematical theorems is grounded fully in logical demonstrations from basic truths of logic; and that the concepts involved in such theorems, and the objects whose existence they imply, are of a purely logical nature. Thus Frege maintained that arithmetic requires no assumptions besides those of logic; that the concept of number is a concept of pure logic; and that numbers themselves are, as he put it, logical objects.

This view of mathematics would not have been possible without a profound transformation of logic that occurred in the late nineteenth century - most especially through the work of Frege. Before that time, actual mathematical reasoning could not be carried out under the recognized logical forms of argument: this circumstance lent considerable plausibility to Immanuel Kant's teaching that mathematical reasoning is not 'purely discursive', but relies upon 'constructions' grounded in intuition. The new logic, however, made it possible to represent standard mathematical reasoning in the form of purely logical derivations - as Frege, on the one hand, and Russell, in collaboration with Whitehead, on the other, undertook to show in detail.

It is now generally held that logicism has been undermined by two developments: first, the discovery that the principles assumed in Frege's major work are inconsistent, and the more or less unsatisfying character (or so it is claimed) of the systems devised to remedy this defect; second, the epoch-making discovery by Kurt Gödel that the 'logic' that would be required for derivability of all mathematical truths can in principle not be 'formalized'. Whether these considerations 'refute' logicism will be considered further below.

1 Preliminary discussion

Although the thesis that arithmetic is a part of logic was stated quite explicitly by Frege (1884, 1893, 1903) and by Russell (1903), it is surprisingly difficult to determine the exact content of this thesis; and, indeed, to determine whether it should be understood to mean the same thing to each of them. For the obvious question presents itself: exactly what are we to understand by 'logic'? How do we tell whether a certain basic assumption, or a certain basic concept, properly belongs to logic, or not? And, since an important part of the thesis is its epistemological aspect - the claim that mathematical knowledge is purely logical in nature - one would naturally expect an account of the grounding of our knowledge of logic. On the first of these questions - how to tell what belongs to logic - Frege has little to say (although he acknowledges the importance of the issue for his philosophy, and does provide a few pregnant suggestions); Russell has some lengthy considerations, which have proved not to be very satisfactory. On the second question, Frege and Russell appear to disagree sharply. Both regard the thesis as being in opposition to Kant; but whereas Frege, in denying that arithmetic is based upon any Kantian 'pure intuition' and maintaining that it is entirely grounded in logic, concludes that Kant was wrong to consider arithmetical propositions synthetic, and holds by contrast that, as propositions of logic, they are analytic (roughly speaking: known to be true by 'analysis' of their meanings alone), Russell denies the latter, and holds that logic itself is synthetic in character:

Kant never doubted for a moment that the propositions of logic are analytic, whereas he rightly perceived that those of mathematics are synthetic. It has since appeared that logic is just as synthetic as all other kinds of truth; but this is a purely philosophical question, which I shall here pass by.

(1903: 457)

Yet even here the issue is not so clearly drawn; for one must still ask whether Frege, in affirming that logic is analytic and not synthetic, and Russell, in affirming the opposite, understood the words 'analytic' and 'synthetic' in the same way - in particular, whether their opposite formulations might none the less have expressed the same opinions. The possibility is lent some weight by the fact that in certain late writings Frege refers to a 'logical source of knowledge' as a distinct knowledge-source (alongside 'sense perception' and 'the geometrical

source of knowledge') ([1924-5] 1979: 267, 278-9). Russell's view that logical truth is synthetic may be essentially identical with Frege's view that knowledge of such truth requires a 'knowledge-source'.

2 Connection with developments in arithmetic, algebra and analysis

The thesis that mathematics is a part of logic could not have been maintained before logic had undergone the deep transformation which took place in the late nineteenth century, most especially through the work of Frege. Before that time, actual mathematical reasoning could not be carried out under the recognized logical forms of argument. The new logic, however, made it possible to represent standard mathematical reasoning in the form of purely logical derivations. It is equally true that no such claim could have been made except for a deep transformation that had taken place in mathematics itself during the nineteenth century; and a brief consideration of this may help to shed some light on a plausible way to construe the logicist thesis.

From the time of Aristotle, it had been usual to regard mathematics as concerned with 'discrete' and 'continuous' quantity: whole numbers on the one hand; continuous magnitudes on the other. In the seventeenth and early eighteenth centuries it was generally supposed that our knowledge of mathematical truths was exclusively concerned with 'relations of ideas', and that - even if these ideas were obtained by some process of abstraction from empirical things - this character explained how it is that the knowledge in question is certain and independent of experienced 'facts' (such a view was common, for instance, to the rationalist Leibniz and the empiricist Hume). Kant had presented arguments to subvert this position: he argued, rather convincingly, not just that logic (as it then existed) was insufficient to warrant the reasoning of mathematicians, but also that such reasoning demanded attention to more than 'ideas', or (rather) concepts - that a kind of mental visualization of the objects falling under mathematical concepts was an indispensable ingredient.

Among mathematicians - although within a fairly small circle - confidence in *either* the Kantian *or* the pre-Kantian traditional view of the nature of mathematical knowledge was disturbed by the development, in the early nineteenth century, of new forms of geometry: these seemed to show that neither mere consideration of 'relations of ideas' (that is, of such concepts as 'point', 'line', 'distance' and so on), nor the appeal to 'intuition', could decide the truth of propositions of geometry (see [Geometry, philosophical issues in](#)). At the same time, attempts were begun to work out new foundations for analysis - that is, the theory of the real (and complex) numbers, functions thereof, and limiting processes - that should be independent of any 'extrinsic' appeal to an 'intuition' of continuous magnitude. Again, in parallel (and partly in connection) with the preceding, systematic procedures were developed for the introduction of 'new objects' - or 'ideal objects' - in several branches of mathematics: for example, 'infinitely distant points' and 'points with imaginary coordinates' in geometry; real and imaginary numbers (alongside the rational numbers) in analysis; and, perhaps most decisively for our subject, 'ideal factors' of algebraic integers in the newly developing subject of algebraic number theory. In this last context in particular, Dedekind achieved fundamental new results with the help of concepts that involved the treatment of *classes* of (already introduced) objects as new objects ('modules', 'ideals', 'orders', 'number fields', among others) which could themselves be made subject both to general mathematical investigation and to actual calculation.

With all of this work in the background, Dedekind was led to examine the foundations of the theory of the whole numbers themselves; and he arrived at the conviction that this theory, together with algebra and analysis which are further derived from it, is 'a part of logic' - in saying which, he tells us, 'I mean to imply that I consider the number-concept entirely independent of the notions or intuitions of space and time, that I consider it *an immediate result from the laws of thought*' ([1888] 1901: 31; emphasis added). The way he develops his theory allows us to conclude that he regards both the concept of number and the theorems of arithmetic, as well as the concepts and theorems of the other mathematical theories he mentions, as resulting from certain fundamental abilities of the mind: the ability to form classes (his word is 'systems') of objects of thought (which themselves then have the status of such objects); to correlate or map such systems onto one another; and to 'create' new objects to 'represent' ones already present ('to relate things to things, to let a thing correspond to a thing, or to represent a thing by a thing'); and he tells us that without this ability, 'no thinking is possible'.

This version of the logicist thesis - namely, that mathematics (or that part of it which is claimed to belong to logic) requires no other source, either of its concepts or of its knowledge, than what is presupposed by all systematic thought - may be taken as applicable to the doctrines of all who have made that claim, whatever epistemological or

metaphysical differences may attach to their further elaboration of what those ‘presuppositions of all thought’ might be.

3 The logicist foundations of arithmetic

The fundamental problem for the logicist theory of the arithmetic of the ordinary whole numbers was to show how (1) the general concept of such numbers, (2) the individual numbers themselves, and (3) the modes of reasoning that suffice for the mathematical demonstrations of their properties, can all be derived from principles of logic. It will be convenient here to consider first the solution offered by Frege (1884), with which Russell’s substantially coincides; a brief comparison with Dedekind’s rather different approach will be made later.

Frege set about creating an entirely systematic and universally applicable formal system codifying all logical concepts and processes: his *Begriffsschrift*, or ‘concept writing’, which he characterized as ‘a formula language...for pure thought’ (1879). He therefore faced the task of making explicit, within the framework of such a formal system, how the cardinal numbers could be defined.

Now for Frege, as for Dedekind, a central logical role was played by a notion of ‘class’ - Dedekind’s ‘system’, characterized by him in the following way:

‘Such a system S ...is completely determined when with respect to every thing it is determined whether it is an element of S or not. The system S is hence the same as the system T ...when every element of S is also an element of T , and every element of T is also an element of S ’

([1888] 1901: 45)

(Compare with ‘extensionality’ in *Set theory, different systems of*.) However, in Frege’s opinion, such notions as that of a system or class, and that of a ‘correspondence’ (‘a thing belongs to a thing’) are ‘not usual in logic and are not reduced [by Dedekind] to acknowledged logical notions’ ([1893] 1964: 4). Frege therefore relies, for the *logical* ground of the notions of class and correspondence, upon the ‘acknowledged’ logical notions of a ‘concept’ (which when it is unary corresponds to a property, and when binary, ternary, ..., to a two-place, three-place, ... relation) and of the ‘extension’ of such a concept (in particular, the extension of a unary concept is a class). The existence of such an extension for every concept, satisfying the principle that two concepts have the same extension if and only if every object falling under either also falls under the other (see Dedekind’s characterization of ‘system’ above), is a fundamental principle of the logic on which Frege bases arithmetic.

To avoid certain verbal complexities, we now follow Russell in formulating the arithmetic notions in terms of classes. Two classes A, B are said to be ‘equinumerous’ if, for some relation R : (1) for every element x of A there is an element y of B such that x has the relation R to y , but to no other element of B ; and (2) for every element y of B there is an element x of A such that x , but no other element of A , has the relation R to y . One says that R is a one-to-one relation between A and B . ‘Equinumerous’, as just defined, is a binary relation that holds between certain classes. Being equinumerous with a given class A is, then, a property which holds of certain classes. In Frege’s usage, the extension of this property [or concept] - and thus, the class of all such classes - is called ‘the number of the class A ’; thus we have succeeded, according to the premises of this whole procedure, in defining the individual numbers as ‘objects’, in purely logical terms. As to the general concept of number, that is now easy: ‘ n is a number’ means that there is some class A such that n is the number of the class A .

A most crucial step remains to be taken. We have defined the concept of number - allegedly at least - in purely logical terms; but this is too wide a concept to serve as the basis for ordinary arithmetic. Indeed, ordinary arithmetic has to do with the non-negative integers, and these, in the present context, are numbers of *finite* sets. But no such restriction appears in our definition of number: we have defined ‘the number of the set A ’ for arbitrary sets - for example, for the set of all the points on a given line. At the same time - what at first will seem a quite unrelated issue - we have yet to show how all arithmetical reasoning can be carried out by purely logical argumentation concerning the number-concept.

The two issues have in fact a deep connection; and it is one of Frege’s great accomplishments (also achieved independently by Dedekind) to have perceived the connection and found in it the simultaneous solution to both of them. A characteristic feature of reasoning in arithmetic is the ‘argument from n to $n + 1$ ’, or the principle of mathematical induction:

If a certain property P holds of the number 0, and if whenever P holds of a number n it holds also of the number $n + 1$, then P holds of all the whole numbers.

(Of course, to formulate this principle at all, one must have defined the process of adding 1 to a given number; but that is a detail, easily accomplished.) The chief problem for placing the argumentation of arithmetic on a ‘purely logical’ footing was to find a logical justification for this principle. But, speaking ‘intuitively’, it is clear that the principle in question cannot apply except to finite numbers (for the property of being finite holds of 0, and holds of $n + 1$ whenever it holds of n ; but of course holds only of finite numbers). This ‘intuitive’ argument can of course have no systematic standing so long as the notion of finite number has not been defined. Frege’s idea, however, was to use this ‘intuitive’ insight as the basis for a purely logical definition of the concept of finite number:

A number x is said to be ‘finite’ if and only if x belongs to every class C such that: (1) 0 belongs to C ; and (2) whenever a number n belongs to C , so does $n + 1$.

With this definition, not only are the finite numbers singled out, but all the remaining concepts of traditional number theory can be introduced, and the reasoning of traditional number theory can be carried out within the framework of Frege’s extended logic.

4 Extension to other mathematical domains

The extension of these results to the arithmetic of the rational, real and complex numbers offers no new fundamental difficulties, after the pioneering work that had already been done by mathematicians (see [Numbers §7](#)).

A difficulty does arise in the theory of functions: namely, the need - first brought to prominent notice by Zermelo (1904), who characterized it as a ‘logical principle’ - for what has come to be called the ‘axiom of choice’; this says, in effect, that if f is a function defined on a domain A , such that, for each x in A , $f(x)$ is a non-empty set, then there is a function g , also defined on A , such that, for each x in A , $g(x)$ is an element of $f(x)$.

An interesting point arises when we come to geometry. Frege never extended the thesis of logicism to geometry; but, in a rather polemical exchange with Hilbert concerning the latter’s path-breaking work on the foundations of geometry (see [Frege 1980](#)), Frege actually came near to formulating the point of view from which this extension can - and, indeed, so far as logicism is viable at all, *should* - be made: namely, one regards geometric theory - or, rather, *a* geometric theory (for there are infinitely many) - as concerned with a species of structure (of any one of which there may be many exemplars), characterized - indeed, defined - by the ‘axioms’ of that geometry; the theorems deduced from those axioms are then demonstrated ‘by pure logic’ to hold of any exemplar of the species so defined. That is the view maintained by Whitehead and Russell in the projected treatment of geometry in their major work, *Principia Mathematica* (1910-13). And it is the view that had already been expressed, in effect, by Dedekind and still earlier by Riemann (see [Stein 1988: 244, 251-3](#)). What is also noteworthy is that Dedekind, in contrast to Frege and to Russell and Whitehead, adopts the same point of view concerning even the system of the whole numbers: that is, he characterizes the *structure* a system must have to serve as that of the whole numbers, in explicit preference to identifying those numbers with any ‘objects’ specified in other than those structural terms.

5 The setbacks; a brief evaluation

The first of the setbacks for the logicist view was the discovery of a contradiction in Frege’s logic (see [Paradoxes of set and property §2](#)), which arises from the assumption, already noted, that every concept has an extension - a very strong form of what has come, in the theory of sets, to be called a ‘comprehension principle’. Frege himself was eventually led to abandon his logicist view of arithmetic, and to suggest instead that arithmetic is not only - as Kant thought - synthetic a priori, but is actually to be grounded in the ‘geometric knowledge-source’ ([\[1924-5\] 1979: 276-80](#)). Others, however - notably Russell ([\[1903\] 1937: introduction](#)), Ramsey (1925) and Carnap (1931) - felt no need to abandon the logicist thesis, but reconciled themselves to a modified view of logic, designed to circumvent the contradiction and still allow the essential modes of reasoning. Most celebrated among logicians as such a modified logic is Russell’s ‘theory of types’; but the axiomatic theory of sets, which has proved most serviceable as a medium for mathematics, could perfectly well also claim to be a form of ‘logic’ (see the characterization quoted from Zermelo, the pioneer of axiomatic set theory, in §4 above).

The second setback is more profound. Frege's own invention of a way of rendering logic purely formal made it possible to apply modes of mathematical reasoning to the formal linguistic structures of logical systems themselves. The undertaking of such application was a fundamental contribution of Hilbert. This process has further revolutionized our way of looking at logic itself. Among the most basic results of this examination are two theorems of Gödel (1930, 1931), one of which tells us that 'elementary logic' - or first-order logic - can be formulated so that every proposition that deserves to be considered logically true can be derived by 'purely formal', or 'mechanical', means (for example, these propositions can be generated by a computer program); whereas the other tells us that no such thing can be achieved for a system - whether a higher-order logic (of which the theory of types is an example), or a first-order axiomatized theory (of which set theory is an example) - that is strong enough to generate the theorems of elementary arithmetic.

With Gödel's theorems there should be mentioned a discovery that preceded them by several years: the Löwenheim-Skolem theorem (see Löwenheim 1915 and, for clearer and more adequate proofs, Skolem 1920, 1923). This theorem implies that any theory axiomatized in first-order logic that has an infinite model (domain that satisfies the axioms) has a countably infinite model (one whose domain can be placed in one-to-one correspondence with the natural numbers). In a very remarkable paper, Skolem (1923) invoked this theorem (and other considerations) to cast grave doubt on the view that set theory could be 'a satisfactory ultimate foundation for mathematics'. Indeed (restricting attention here to the implications of the Löwenheim-Skolem theorem), it is, for example, a fundamental proposition of classical analysis that the set of all the real numbers is uncountably infinite. The existence of a model for the allegedly foundational theory containing only countably many elements seems, then, *prima facie* to show that the foundational attempt has failed: that, in particular, the assertion within the theory that a certain set is uncountable has not captured what ought to be the genuine meaning of the term 'uncountable'. The issue is both deep and subtle; here, it must suffice to remark that both Gödel's theorems and the Löwenheim-Skolem theorem lead to the notion of nonstandard models - the theory of which has since become an important part of logic (see [Löwenheim-Skolem theorems and nonstandard models](#)).

These results may be summed up by saying that *there is not, and in principle cannot be, a 'completely formalized' logic within which all mathematically statable questions are correctly answerable*. Whether one chooses to call this a refutation of logicism, or an unexpected result about the 'uncompletable' nature of logic itself, the result is unquestionably a deep and interesting one simultaneously concerning mathematics and logic; and, for its discovery, a great debt is owed to those who advanced the thesis of logicism.

See also: [Arithmetic, philosophical issues in §4](#); [Hilbert's programme and formalism](#); [Intuitionism §1](#); [Logical and mathematical terms, glossary of](#)

HOWARD STEIN

References and further reading

- Carnap, R.** (1931) 'Die logizistische Grundlegung der Mathematik', *Erkenntnis* 2: 91-105; trans. E. Putnam and G. Massey, 'The Logician Foundations of Mathematics', in P. Benacerraf and H. Putnam (eds) *Philosophy of Mathematics: Selected Readings*, Cambridge: Cambridge University Press, 2nd edn, 1983, 41-52. (A classic account of logicism. Referred to in §5.)
- Carnap, R.** (1939) *Foundations of Logic and Mathematics*, Chicago, IL: University of Chicago Press. (A lucid account of Carnap's view of the relation of logic to the analysis of language, and of how mathematics - a part of the purely logical aspect of language - acquires relevance for the formulation of propositions of empirical science.)
- Dedekind, R.** (1888) *Was sind und was sollen die Zahlen?*, Braunschweig: Vieweg; repr. in *Gesammelte mathematische Werke (Collected Mathematical Works)*, ed. R. Fricke, E. Noether and Ö. Ore, Braunschweig: Vieweg, 1932, vol. 3, 335-90; trans. W.W. Beman (1901), 'The Nature and Meaning of Numbers', in *Essays on the Theory of Numbers*, New York: Dover, 1963. (Dedekind's theory of the natural numbers. Referred to in §§2-4.)
- Frege, G.** (1879) *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*, Halle: Nebert; repr. Hildesheim: Olms, 1964; trans. S. Bauer-Mengelberg, 'Begriffsschrift, a Formula Language, Modelled Upon That of Arithmetic, for Pure Thought', in van Heijenoort (1967), 5-82. (Includes the first formulation of Frege's system of logic. Referred to in §3.)
- Frege, G.** (1884) *Die Grundlagen der Arithmetik: eine logisch-mathematische Untersuchung über den Begriff der*

- Zahl*, Breslau: Koebner; repr. and trans. J.L. Austin, *The Foundations of Arithmetic: A Logico-Mathematical Enquiry into the Concept of Number*, Evanston, IL: Northwestern University Press, 2nd edn, 1980. (An extended philosophical essay which culminates in Frege's definitions of number and finite number, and his analysis of mathematical induction; referred to in §3.)
- Frege, G.** (1893, 1903) *Grundgesetze der Arithmetik: begriffsschriftlich abgeleitet*, Jena: Pohle, 2 vols; repr. Hildesheim: Olms, 1966; part 1 of vol. 1 trans. M. Furth, *Basic Laws of Arithmetic: An Exposition of the System*, Berkeley, CA: University of California Press, 1964. (The first two volumes of Frege's projected complete systematic treatment of arithmetic, including the theory of the real numbers, on an explicitly logical foundation. In consequence of Russell's discovery of a contradiction in the system, the third volume was never written. Referred to in §§1, 3.)
- Frege, G.** (1924-5) 'Erkenntnisquellen der Mathematik und der mathematischen Naturwissenschaften', 'Zahlen und Arithmetik' and 'Versuch einer neuen Begründung für die Arithmetik', in *Nachgelassene Schriften und Wissenschaftlicher Briefwechsel*, vol. 1, ed. H. Hermes, F. Kambartel and F. Kaulbach, Hamburg: Meiner, 1969; trans. P. Long and R. White, 'Sources of Knowledge of Mathematics and the Mathematical Natural Sciences', 'Numbers and Arithmetic' and 'A New Attempt at a Foundation for Arithmetic', in *Posthumous Writings*, ed. H. Hermes, F. Kambartel and F. Kaulbach, Chicago, IL: University of Chicago Press, 1979. (Three very late brief manuscripts, giving Frege's final thoughts on the foundations of mathematics. Referred to in §§1, 5.)
- Frege, G.** (1980) *Philosophical and Mathematical Correspondence*, ed. B. McGuinness, Chicago, IL: University of Chicago Press, 34-48. (Frege's correspondence with Hilbert on the latter's *Foundations of Geometry*. Referred to in §4.)
- Gödel, K.** (1930) 'Die Vollständigkeit der Axiome des logischen Funktionenkalküls', *Monatshefte für Mathematik und Physik* 37: 349-60; trans. S. Bauer-Mengelberg, 'The Completeness of the Axioms of the Functional Calculus of Logic', in van Heijenoort (1967), 582-91. (Gödel's proof of the completeness of first-order logic; referred to in §5.)
- Gödel, K.** (1931) 'Über formal unentscheidbare Sätze der *Principia Mathematica* und verwandter Systeme I', *Monatshefte für Mathematik und Physik* 38: 173-98; trans. J. van Heijenoort, 'On Formally Undecidable Propositions of *Principia Mathematica* and Related Systems', in van Heijenoort (1967), 592-617. (Gödel's celebrated incompleteness theorems, the first of which is referred to in §5.)
- Heijenoort, J. van** (ed.) (1967) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press. (Includes translations of several of the works discussed in this article, with introductions.)
- Löwenheim, L.** (1915) 'Über Möglichkeiten im Relativkalkül', *Mathematische Annalen* 76: 447-70; trans. S. Bauer-Mengelberg, 'On Possibilities in the Calculus of Relatives', in van Heijenoort (1967), 232-51. (Löwenheim's original proof - difficult, and couched in now unfamiliar notation - of the Löwenheim-Skolem theorem; van Heijenoort's introduction is very helpful. Referred to in §5.)
- Ramsey, F.P.** (1925) The Foundations of Mathematics, *Proceedings of the London Mathematical Society*, series 2, 25(5): 338-84; repr. in *The Foundations of Mathematics and Other Logical Essays*, ed. R.B. Braithwaite, London: Routledge, 1931. (Referred to in §5. The posthumous collection, of both previously published and unpublished works, in which this essay appears includes important discussions of the theory of types, the axiom of choice and (a point not discussed in the text above) Russell and Whitehead's axiom of infinity - needed in their version of a logicist theory in order to obtain a complete system of arithmetic.)
- Russell, B.A.W.** (1903) *The Principles of Mathematics*, Cambridge: Cambridge University Press; 2nd edn, London: Allen & Unwin, 1937; repr. London: Routledge, 1992. (A historically important work, of substantial difficulty; quoted in §1.)
- Skolem, T.** (1920) 'Logisch-kombinatorische Untersuchungen über die Erfüllbarkeit oder Beweisbarkeit mathematischer Sätze nebst einem Theorem über dichte Mengen', *Videnskapsselskapets skrifter, I. Matematisk-naturvidenskabelig klasse 4*; §1 trans. S. Bauer-Mengelberg, 'Logico-Combinatorial Investigations in the Satisfiability or Provability of Mathematical Propositions: A Simplified Proof of a Theorem by L. Löwenheim and Generalizations of the Theorem', in van Heijenoort (1967), 254-63. (Skolem's more satisfactory proof, and strengthening, of Löwenheim's theorem on countable models; referred to in §5.)
- Skolem, T.** (1923) 'Einige Bemerkungen zur axiomatischen Begründung der Mengenlehre', in *Matematikerkongressen i Helsingfors den 4-7 Juli 1922, Den femte skandinaviska matematikerkongressen, Redogörelse*, Helsinki: Akademiska Bokhandeln, 217-32; trans. S. Bauer-Mengelberg, 'Some Remarks on Axiomatized Set Theory', in van Heijenoort (1967), 291-301. (A brief, remarkably clear and very deep discussion of a series of issues concerning the foundations of mathematics. Referred to in §5.)

- Stein, H.** (1988) 'Logos, Logic, and Logistiké: Some Philosophical Remarks on the Nineteenth-Century Transformation of Mathematics', in W. Aspray and P. Kitcher (eds) *History and Philosophy of Modern Mathematics*, Minneapolis, MN: University of Minnesota Press, 238-59. (Includes a discussion of some of the nineteenth-century developments that form the mathematical background to the logicist thesis referred to in §2; and gives an indication of the views of Riemann and Dedekind on geometry, referred to in §4.)
- Whitehead, A.N. and Russell, B.A.W.** (1910-13) *Principia Mathematica*, Cambridge: Cambridge University Press, 3 vols; 2nd edn, 1925-7. (Referred to in §4.)
- Zermelo, E.** (1904) 'Beweis, daß jede Menge wohlgeordnet werden kann', *Mathematische Annalen* 59: 514-16; trans. S. Bauer-Mengelberg, 'Proof that Every Set can be Well-Ordered', in van Heijenoort (1967), 139-41. (The article that brought prominently to the attention of mathematicians the axiom of choice - here 'the assumption that coverings exist'; referred to in §4.)

Löwenheim-Skolem theorems and nonstandard models

Sometimes we specify a structure by giving a description and counting anything that satisfies the description as just another model of it. But at other times we start from a conception we try to articulate, and then our articulation may fail to pin down what we had in mind. Sets seem to have had such a fate. For millennia sets lay fallow in logic, but when cultivated by mathematics in the nineteenth century, they seemed to bear both a foundation and a theory of the infinite. The paradoxes of set theory seemed to threaten this promise. With an eye to proving freedom from paradox, versions of set theory were articulated rigorously. But around 1920, Löwenheim and Skolem proved that no such formalized set theory can come out true only in the hugely infinite world it seemed to reveal, for if it is true in such a world, it will also be true in a world of the smallest infinite size. (Versions of this remain true even if we augment the standard expressive devices used to formalize set theory.) But then, Skolem inferred, we cannot articulate sets determinately enough for them to constitute a firm foundation for mathematics.

1 Background

To see how a model could fail to be standard, one first needs some idea of what a model is supposed to be. The basic distinction here is that between words and the world, signs and things. If the extension (denotation, reference) of each sign in a sentence or collection of sentences is held utterly fixed, and no re-interpretation is allowed at all, then what that sentence or collection commits us to should also be fixed. In that case there is little point in trying to distinguish various ways those sentences might come out true, that is, various models for them. If, in contrast, the extensions of none of the signs are in any way constrained, then just about anything could be gerrymandered into a way of making those sentences come out true. In this case too it seems unlikely that there will be interesting distinctions among the models. So the interest in models for a collection of sentences depends on holding the extensions for some but not all of the signs in those sentences fixed. (See [Model theory](#).)

In modern algebra, a group is defined as a non-empty set G and a function $\circ$ such that:

- (a) whenever a, b are in G , so is $a \circ b$;
- (b) whenever a, b, c are in G , then $a \circ (b \circ c) = (a \circ b) \circ c$;
- (c) there is a member e of G such that $a \circ e = e \circ a = a$ for each a in G ; and
- (d) for each a in G there is a member a^{-1} of G such that $a \circ a^{-1} = a^{-1} \circ a = e$.

The integers under addition constitute a group. The one-to-one correspondences between a set and itself also form a group under composition of those correspondences. Note that any model for the axioms is a group; there is no question of any groups being somehow standard, and others not. (On the other hand, because addition of integers commutes, that is, $a + b = b + a$, while composition of one-to-one correspondences need not, composition of correspondences illustrates groups in general better than addition of integers.)

If we put the second axiom of groups into the notation of quantification theory, we get

$$(\forall a, b, c)(a, b, c \in G \rightarrow a \circ (b \circ c) = (a \circ b) \circ c).$$

Any choice of a non-empty set for G and a binary function on G for $\circ$ is a candidate for a structure or interpretation making this axiom true. In this way, the extensions of the signs ' G ' and ' $\circ$ ' are allowed a huge range of variation. But those of ' $\forall$ ', ' $\in$ ', ' $\rightarrow$ ' and ' $=$ ' are held fixed; they must be taken as universal quantification, membership, the conditional, and identity. This mixture of variation and constancy is part of what gives groups an interesting structure.

Modern algebra was a nineteenth-century development, and groups are an artefact of that growth. There was no favoured or intended group to begin with; it is instead the general group structure that proved to be of interest. So there is no distinction between standard and nonstandard models of the group axioms.

We think of groups as *given* by their axioms; they have no life prior to their characterization that would distinguish intended models of that description from unintended ones. But sets probably began life among us as extensions of predicates. On the one side there are predicates like 'is a lion'. On the other, there are the lions. One way words and things link up is that all and only the lions are members of the extension of the predicate 'is a lion'. We take the extension of a predicate as the collection, set or class of all and only the things of which that predicate is true.

Disguised as extensions sets led fairly inactive, boring lives in logic for ages.

But around the turn of the nineteenth century, reliance on geometric intuition led analysis, the theory growing out of differential and integral calculus, into paradox; intuition is not always trustworthy on infinitary processes. So a movement arose in nineteenth-century mathematics to replace geometry with arithmetic. We had learned in analytic geometry to individuate a point in the plane by an ordered pair of numbers, the point's abscissa and ordinate. Now we could rethink a literal curve passing (for example, continuously) through such points as some sort of arithmetical law, or even equation, relating the abscissas of such points to their ordinates. This rethinking of analysis is called its arithmetization.

The arithmetization of analysis focused attention on the kind of numbers used as coordinates of points on the plane. Nowadays we call these real numbers; they are the numbers used for lengths, temperatures and many other natural quantities, and given by decimals. Arithmetizing analysis calls for a clear view of the reals and their laws. To those with a taste for abstraction or essence, it even raises the question of what a real number is. Later on in the nineteenth century, Richard Dedekind showed how to answer this question by taking reals as (roughly) certain sets of rational numbers. (One can think of rationals as quotients of integers in lowest terms.) The details to one side, the idea is to rethink reals as certain constructions out of rationals using sets to do the constructing. Such constructions can be shown to satisfy the laws of reals needed to do analysis (see [Dedekind, J.W.R.](#)).

At that point we seem to be on to something. Reals are constructed from rationals using sets. It is a lot easier to construct rationals from integers (positive, negative and zero) using sets and to construct integers from natural numbers (non-negative whole numbers) using sets. Then Frege and Cantor seem independently to have got the idea of constructing natural numbers out of sets alone (at least that is how we in hindsight might put it). Sets seem to be becoming the basic stuff of mathematics, so if these sets are the same old extensions of logic, then perhaps, as Frege claimed, mathematics is reducible to logic (see [Cantor, G.](#); [Frege, G.](#)).

The basic idea was to begin with the as-many-as relation. One set A has as many members as another set B precisely in case there is a binary relation R such that (1) for each $a \in A$ there is exactly one $b \in B$ such that aRb ; (2) R matches different members of A with different members of B ; and (3) for each $b \in B$ there is an $a \in A$ such that aRb . Beyond set-membership, only notions of logic (all, some, identity, and so on) are needed for this exposition of as-many-as. But with it, Cantor and Frege could take the number N_A of members of a set A as the set of all sets B with as many members as A , and a number as the number of members of some set (see [Logicism](#)).

But there is more. Mathematics had always looked to be about things like numbers, of which there are infinitely many, and triangles, of which there are infinitely many and in each of which there are infinitely many points. But mathematics lacked a good general account of the infinite *per se*. Dedekind provided one; a set is infinite precisely in case it has exactly as many members as one of its proper subsets (that is, a set included in but not exhausting the original set). Dedekind's account might have been only a curiosity had Cantor not proved that for each set, even infinite, there is a strictly larger set. At a single stroke, Cantor revealed a wealth of different infinite sizes, and thus a realm of transfinite arithmetic, which Hilbert called Cantor's paradise. Set theory looked to promise both a foundation for mathematics and a whole new mathematics of the infinite (see [Set theory](#); [Cantor's theorem](#); [Continuum hypothesis](#)).

But there were serpents in paradise. Most philosophers have seen Russell's paradox. Cantor's starts from a predicate like 'is a thing' which is true of everything. Every set would have to be a subset of U , the extension of this predicate, and so be no larger than U . But by Cantor's result cited in the previous paragraph, there is a set strictly larger than U . This contradiction together with Russell's and a third due to Burali-Forti constitute the paradoxes of set theory, and around the turn of the twentieth century some regarded these paradoxes as a crisis in the foundation of mathematics (see [Paradoxes of set and property](#)).

One response to this crisis was to formulate the mathematics at issue with unprecedented rigour. We would be utterly explicit about exactly which signs (that is, marks) may be written, how these marks may be concatenated, and how such strings may be manipulated. This idea of a formal system had been implicit in Frege, but it acquires a new prominence in Hilbert's programme for responding to the crisis in the foundation of mathematics. For once we have such a formal system we can ask whether anything with the shape of a contradiction can be cranked out by its rules of manipulation. Here we might prescind from all questions about how these marks might be

interpreted and focus instead only on which finite combinations of them the rules for their manipulation permit. This latter inquiry sounds like very elementary mathematics indeed. Hilbert's programme might look to show how to use only the most concrete and trustworthy mathematics, combinatorics, to prove the consistency of, and so to that extent legitimate trust in, the most basic (and so most powerful) set theory, at least as formalized (see [Formal languages and systems](#); [Hilbert's programme and formalism](#)).

But now we have come back to a distinction between words and the world. On the one side, in the world, we have the sets. These are not just any old sets; they are to be a foundation for all of mathematics and to embody a new mathematics of the infinite. We have expectations of these sets; there are, for example, to be enormously many sets, many more than there are natural numbers or even real numbers. So much was promised by Cantor's theorem long before set theory got formalized, or even axiomatized. On the other side, in language, we have formalized set theory. To formalize it means to give its primitive marks and the rules for their manipulation just in terms of their shapes without a thought for how these marks might be assigned things in the world as extensions or interpretations. We want that purely formal presentation in hopes of as concrete and combinatorial a consistency proof for the system as possible. One might have other aims in formalizing, but Hilbert wanted a consistency proof as a *laissez-passer* back into Cantor's paradise. But having made this distinction between words and the world, one then can see the possibility of the words having an interpretation in the world quite different from what was intended.

2 One proof of the Löwenheim-Skolem theorem

In 1915 Leopold Löwenheim, a secondary school teacher, published a claim that a formula in the calculus of relatives that comes out true under an interpretation in an infinite domain also comes out true under an interpretation in a domain of the smallest infinite size. This size is that of the natural numbers, and domains of that size are called countably infinite. We might now describe the formulas in which Löwenheim was interested as those in which only the truth functional connectives and the quantifiers are interpreted uniformly across domains, that is, non-empty sets; the other signs, that is, constants, function letters and predicate letters, may be interpreted by any member of, or function or relation (of suitable polyadicity) on, any such domain. The consensus nowadays is that Löwenheim failed to prove his claim.

But after the First World War, the Norwegian logician Thoralf Skolem gave two rather different but equally correct proofs of Löwenheim's claim; this is the origin of the Löwenheim-Skolem theorem. Nowadays, one of these arguments is usually presented with a proof of the completeness of quantification theory (first-order logic). To show completeness, we start from a valid formula, that is, one true under all interpretations in all non-empty domains, and we must show that it has a proof by the rules of quantification theory. Equivalently by contraposition, we must show that a formula without such a proof fails in some interpretation. For this it suffices to show that any formula that yields no explicit contradiction by the rules is true in some interpretation. We can in fact show that the domain of such an interpretation need be no larger than countably infinite. But any formula true in some hugely infinite domain cannot yield a contradiction by the rules, and so by our result is true in an at most countably infinite domain.

Skolem not only proved Löwenheim's claim: he also strengthened it. For example, the claim holds not just for single formulas, but for countably infinite sets of formulas. Now remember our precisely formalized set theory. As standardly presented, this can be taken as a countably infinite set of sentences in which, over and above logic (which may be taken to include identity), there is only a single binary predicate (intended for the membership relation). Replace this binary predicate by a dyadic predicate letter allowed to be interpreted by any two-place relation on any non-empty domain. We get a set of formulas to which Skolem's strengthening of Löwenheim's result applies. Call this set of formulas the logical skeleton of set theory. It has lost its meat, the real extension of the predicate originally intended for the membership relation. It can be taken as the set of formulas that would be true were the predicate interpreted as intended, not just the theorems of some axiomatization of set theory. If set theory as standardly presented is, interpreted as intended, true, then because of Cantor's result on different infinite sizes, this set of formulas is true in a hugely infinite domain, namely, the world; and so by the Löwenheim-Skolem theorem, it is also true in an at most countably infinite domain. Indeed, since a countable model always has an isomorphic copy in the natural numbers, that countable model can be taken as one on the natural numbers. Hence the logical skeleton of set theory cannot come out true only when it is interpreted as intended in set theory. One

might begin to worry whether the laws of sets pin them down determinately enough for them to be a firm foundation for mathematics.

But raising this worry, as we have, with a spotlight on the logical skeleton of set theory suggests a way to assuage it. Surely parallel evolution could result in two very different animals whose skeletons are none the less indistinguishable. If so, an animal's skeleton need not individuate it. Similarly, the wealth of sets comes not from the logical form of (first-order) truths about membership, but rather from the nature of membership. For by the Löwenheim-Skolem theorem, the logical forms of the first-order truths of set theory come out true in a world vastly smaller than the real world of sets. So those forms alone fail to pin down the full nature of the membership relation, which is what makes the real world so big. That nature cannot be isolated just by the logical form of first order truths about it. Note in this connection that the proof of the Löwenheim-Skolem theorem sketched above makes no use of the huge model it supposes except to infer simple consistency; it constructs the small model from the merely consistent theory alone. Perhaps it is not surprising that with the meat thrown away, bones no longer carry its tang.

3 Another proof of the theorem and Skolem's paradox

But Skolem also proved the theorem in another way that makes much more intimate use of the huge original model. To get the gist, we need the idea of a Skolem function, which is another way of taking a quantifier in a model. A formula, perhaps with free variables, is in prenex form when all its truth functional connectives lie in the scopes of all its quantifiers. There are mechanical routines for finding prenex equivalents of formulas. Take prenex forms for each of the formulas in the logical skeleton T of set theory. Let these formulas be interpreted as intended in a huge domain D of sets on which E is the membership relation, the intended extension of the binary predicate letter that is the only extra-logical sign in T . (There may be a problem here. For model theory to be a theory, the domains of models should be objects we know how to handle. For that reason, they are usually required to be sets. But the standard response to paradoxes like Cantor's is nowadays to deny that there is a set of all sets. So if D is to be treated by the laws of all sets (as it is), then it had better not exhaust the sets, in which case E cannot be *the* membership relation.) Next we glance at orderings of D . A linear ordering of D is a two-place relation that arranges its elements as in a straight line; picture the relation of being-to-the-right-of. A well-ordering of D is a linear ordering R of D such that every non-empty subset of D has an R -least member, which then by linearity is unique. Being-to-the-right-of is not a well-ordering of the points on a line. But an equivalent of the axiom of choice of set theory says there is a well-ordering R of D (see **Axiom of choice**). Pick an arbitrary member d of D for don't-care cases. Let

$$(\forall x)(\exists y)M(x, y, z)$$

be one of our prenex forms, where M is a quantifier-free truth functional compound in which only x , y and z are variables. Skolem functions come from existential quantifiers in such prenexes. The existential quantifier binding y lies in the scope of one universal quantifier binding x and in a formula in which z occurs free. So the Skolem function for this quantifier is a two-place function f of the arguments x and z . For any members x and z of D , the value of f at x and z is the R -least member a of D such that $M(x, a, z)$ if there is any such a at all, but d otherwise (the don't-care case). Skolem functions are a little more complex in general, but that is the basic idea.

Take one prenex form for each formula in T . That is a countable infinity of forms, each of which has only finitely many existential quantifiers in its prenex. So there is only a countable infinity of Skolem functions. Pick any countable subset A of D , but toss d into it. Let A_0 be A , and for each natural number n , let A_{n+1} be A_n together with all the values of all the Skolem functions on members of A_n . We can show that for each n , A_n is countable, and thus that $D' = A_0 \cup A_1 \cup \dots$ is also countable. D' is also closed under all Skolem functions; that is, if f is a k -adic Skolem function and $d_1, \dots, d_k \in D'$, then $f(d_1, \dots, d_k)$ is also in D' . This last is crucial; it comes from the fact that any prenex of any formula is finitely long, which comes from the fact that formulas are of finite size. Now let E' be E cut down to D' ; that is, $aE'b$ if and only if aEb , and a, b are both in D' . The relation E' on the domain D' constitutes our new small model of T ; it is called the Skolem hull of our huge old model.

A definition of satisfaction converts the extension E' of T 's only extra-logical sign into extensions in D' for each of its formulas. (See **Tarski's definition of truth**.) It is only for members of D' that there is much point in asking whether they do or do not belong to these extensions. Satisfaction also converts E into extensions in D for

formulas of T , and we can compare how these extensions treat objects from the overlap of their domains, namely D' . Let F be a formula of T in which the variables $x_1, \dots, x_k$ occur free and let $d_1, \dots, d_k$ be members of D' . Then we can show that $d_1, \dots, d_k$ satisfy F as interpreted by E in D if and only if they satisfy it as interpreted by E' in D' . The two interpretations agree where both make sense. This includes all sentences of T , so the same sentences of T are true in both. All the truths of set theory were to hold in the big interpretation, so they hold in the smaller one, even the formalization of Cantor's theorem.

This time around E' has been carved directly out of E . All that has been pared away is whatever in D was unnecessary to make the formulas of T true when interpreted by E . But that turns out to be almost all of D . In particular, although a formula of T was intended under E to say that there are more than countably many things (in D), and though E' has enough of E for both to satisfy the same formulas in D' , that formula intended to assert uncountability is true in D' even though D' is countable.

In 1922 Skolem wrote:

axiomatizing set theory leads to a relativity of set theoretic notions, and this relativity is inseparably bound up with every thoroughgoing axiomatization.... In order to obtain something absolutely non-denumerable, we would have to have either an absolutely non-denumerably infinite number of axioms or an axiom that could yield an absolutely non-denumerable number of first-order propositions. But this would in all cases lead to a circular introduction of the higher infinities; that is, on an axiomatic basis higher infinities exist only in a relative sense.

(Skolem 1922)

These words are the origin of what has come to be called Skolem's paradox. It is pretty cut and dried what Russell's paradox is, but not Skolem's. Just what distinction between absoluteness and relativity did Skolem intend? Was it to be the same for notions as for existence? How does it line up with the countable/uncountable distinction? Do his results rule axiomatic set theory out as a foundation for mathematics?

These questions might lead us in many directions, but we have space only to gesture in one. When, then, might formulas capture, express or pin down a notion? One fairly evident idea is that they do so just in case all and only the models for those formulas exemplify the notion. In this way, the axioms for a group pin down the notion of a group because the groups are, not surprisingly, the models for those axioms. Perhaps formulas catch a notion absolutely if all and only the models for those formulas exemplify the notion. Then we note that uncountability is exemplified by some but not all models for T , so T does not capture uncountability absolutely but only, Skolem may have meant, at best relatively. (Maybe, but now there seem to be many things he might have meant.) Any first-order theory with uncountable models also has countable models, so no first-order theory guarantees that its models are countable, and thus satisfaction of a first-order theory does not guarantee the existence of an uncountable infinity. But satisfaction of the axioms of a group does guarantee the existence of a group. Some such contrast may be what Skolem meant by calling the existence of uncountable infinities relative.

4 One try at a solution to the paradox

Skolem seems to have had it in for set theory. But maybe the Löwenheim-Skolem theorem reveals not so much flaws in axiomatic set theory as limitations in the expressive powers of first-order logical notation. Skolem's words above hint that only a somehow question begging uncountable infinity of notations could guarantee uncountability in models of those notations. The circle is too faint. Is it vicious to use ink to describe ink, or to describe finitude in finitely many words? The hint seems suspect. Add to the notation a new quantifier ' Ux ' such that ' $(Ux)Fx$ ' is true only in a domain with uncountably many members that satisfy F , and count as models only interpretations that verify ' $(Ux)(x = x)$ '. Would limiting ourselves to such models be any more circular than limiting ourselves to models without non-self-identical members?

Second-order quantification is quantification into the positions of predicates, rather than singular terms, in sentences. (See [Second-order logic, philosophical issues in.](#)) Such quantification might seem less obtrusive than uncountable individual quantification, but it also lets us say that there are uncountably many things, though framing the claim calls for a bit more ingenuity. But if the domains of models for formulas with quantifiers of either sort must be sets, then discourse with either sort of quantifier is subject to generalized versions of

Löwenheim-Skolem. Formal languages are usually distinguished by choices among countable infinities of individual constants, predicate letters and function signs. If so, there are only as many such languages as there are reals, and though there are lots of reals, they make up a set. So there are only that many theories, and thus satisfiable theories, too. To each such satisfiable theory assign the least cardinal that is the size of the domain of a model of the theory; all sets have cardinals, so if domains must be sets, this assignment works. Then (by the axiom of replacement) there is a set of all these least cardinals, and so a smallest cardinal greater than all of them. Whether we allow uncountable individual or second-order quantification, this cardinal is to those theories as the cardinal next after that of the natural numbers is to first-order theories according to the Löwenheim-Skolem theorem. For each is a size smaller than that of the intended world of set theory, yet each is a size larger than that of a world in which set theory comes out true. So long as the domains of models must be sets and languages distinguished by choices from a set of primitives, some such result seems inevitable whatever expressive devices we allow our languages; but if the domains need not be sets, it is unclear what they might be and how we might handle them.

See also: [Logical and mathematical terms, glossary of](#)

W.D. HART

References and further reading

- Chang, C.C. and Keisler, H.J.** (1973) *Model Theory*, Amsterdam: North Holland, 3rd edn, 1990, 141-2. (The authors present a quick but contemporary version of Skolem functions and hulls.)
- Fraassen, B.C. van** (1971) *Formal Semantics and Logic*, New York: Macmillan, 122-5. (The author presents a tidy exposition of Skolem hulls.)
- Hart, W.D.** (1970) 'Skolem's Promises and Paradoxes', *Journal of Philosophy* (1970): 98-109. (An exposition and discussion of the paradox from a Platonist perspective.)
- Quine, W.V.** (1976) 'Ontological Reduction and the World of Numbers', in *The Ways of Paradox and Other Essays*, revised edn, Cambridge, MA: Harvard University Press, esp. 212-20. (Unlike Skolem, Quine worries about whether the theorem shows that Pythagoras was right and only numbers exist.)
- Skolem, T.** (1920) 'Logico-Combinatorial Investigations in the Satisfiability or Provability of Mathematical Propositions', in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967, 252-63. (Skolem proves the version of the theorem using the axiom of choice.)
- (1922) 'Some Remarks on Axiomatized Set Theory', in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967, 290-310. (Skolem proves the version of the theorem without the axiom of choice. This paper contains his original presentation of his paradox.)
- Tarski, A. and Vaught, R.** (1956) 'Arithmetical Extensions of Relational Systems', *Compositio Mathematica* (1956): 81-102. (The authors set out the basics of the contemporary understanding of how Skolem functions and hulls work.)

Many-valued logics

Many-valued logics may be distinguished from classical logic on purely semantic grounds. One of the simplifying assumptions on which classical logic is based is the thesis of bivalence, which states that there are only two truth-values - true and false - and every sentence must be one or the other. Many-valued logics reject the thesis of bivalence and permit more than two truth-values.

1 Motivations

Bivalence, in conjunction with some plausible assumptions about the nature of time, seems to lead to determinism. It seems plausible that no matter what I do now, I cannot change what was the case yesterday. This fact suggests that statements about the past are not contingent in the way most of us feel that statements about the future are. Suppose we want to know whether Charles will go sailing tomorrow, Saturday. We think Charles is a free agent, and his going sailing tomorrow is dependent on how he feels, his view of the weather, and other such factors. Consider the sentence ‘Yesterday it was the case that Charles will go sailing on Saturday’. Since yesterday is past, the truth-value of this sentence is fixed, even if we do not now know what it is. If the sentence is true, then it must be the case that Charles is going to go sailing tomorrow; and if the sentence is false, then it must be the case that Charles is not going to go sailing tomorrow. Either way, it seems that whether or not Charles is going sailing tomorrow is already determined, so the notion of future contingency is an illusion.

One way of escaping this argument for determinism is to reject the thesis of bivalence. The most obvious first approach would be to insist on three truth-values: ‘true’, ‘false’ and ‘indeterminate’. This problem was discussed by Aristotle (in *De Interpretatione*), although his position on the issue of bivalence remains a matter of debate. Indeterminists, such as the Epicureans, rejected the thesis of bivalence; determinists, such as the Stoics, advocated it. In the modern era, the Polish logician Łukasiewicz developed several many-valued logics as a means to escape the determinist argument (see [Łukasiewicz, J. §3](#)).

Many other motivations for many-valued logics have been advanced. In every case, there is much debate over the extent to which many-valued logic solves the problems which motivated its advocacy.

Another ancient motivation for expanding the stock of truth-values was the consideration of modalities. Some sentences, for example, ‘Two plus two is four’, seem to be not just true, but *necessarily* true. ‘The sun is shining today’, however, may well be true, but is not necessarily true since we can imagine an alternative. Such considerations suggest having at least four truth-values: ‘necessarily true’, ‘contingently true’, ‘contingently false’, ‘necessarily false’. However, it has been shown that most important systems of modal logic do not correspond to any finite-valued semantics.

Another common motivation for many-valued logics is the consideration of uncertainty. Uncertainty from many sources has been considered: from the incompleteness of human knowledge; from the lack of termination of machine computations; from the conflicting judgment of expert opinion; from true statistical fluctuations in nature; and from accumulating error. A major problem is that uncertainties do not combine in a truth-functional way, but many-valued logics are generally truth-functional.

Many attributions which we commonly make every day (for example, ‘Charles is ugly’) seem to be vague, that is, subject to degree, rather than being black or white (see [Vagueness](#)). Investigations of many-valued logics have been motivated by attempts to deal with inherently vague aspects of natural language. Considerations of vagueness have led to interest in so-called ‘fuzzy logic’ in computing science and electrical engineering (see [Fuzzy logic](#)).

Many-valued logics have been advocated as a means of avoiding various paradoxes, including the usual semantic paradoxes (see [Semantic paradoxes and theories of truth](#)), such as the liar paradox, as well as paradoxes associated with the comprehension axiom of set theory (see [Paradoxes of set and property](#)). For the semantic paradoxes, unless the syntax of the language is unnaturally restricted so that crucial aspects of the semantics cannot even be expressed, versions of the paradoxes simply reappear. For example, suppose we have three truth-values, and consider the sentence ‘This sentence is either false or has indeterminate truth-value’. If its value is false, then what it says is true; if its value is indeterminate then what it says is true; but if its value is true, then it is either false or indeterminate. Similarly, the paradoxes of adopting the unrestricted comprehension axiom in axiomatic set theory

cannot be avoided solely by adopting a many-valued logic.

2 An example

As a simple example, suppose that instead of the classical semantic range of just ‘true’ and ‘false’, we use the following.

T:	logically true
t:	empirically true
U:	undeterminable
f:	empirically false
F:	logically false

The values T and t are said to be ‘designated’ values, or ways of being true. The other values are said to be ‘undesigned’. We could give semantic characterizations for a great number of connectives, but here are just a few examples.

A	$J_T(A)$	$J_t(A)$	$J_U(A)$	$J_f(A)$	$J_F(A)$	$\sim A$	$\neg A$
T	T	F	F	F	F	F	F
t	F	T	F	F	F	f	F
U	F	F	T	F	F	U	T
f	F	F	F	T	F	t	T
F	F	F	F	F	T	T	T

$\supset$	T	t	U	f	F
T	T	t	U	f	F
t	t	t	U	f	f
U	U	U	U	U	U
f	U	U	U	U	U
F	U	U	U	U	U

$\rightarrow$	T	t	U	f	F
T	T	t	f	f	F
t	T	t	t	f	f
U	T	t	t	t	t
f	T	t	t	t	t
F	T	T	T	T	T

$\langle \rangle \langle eq \rangle$

The J_i connectives have the property that $J_i(A)$ has the value T if and only if A has the value i . They are known as ‘autodescriptive’ operators; they allow the expression in the syntax of the underlying semantic motivation. While $\neg$ corresponds to ‘is not designated’, the connective $\sim$ is the more usual complement negation. The conditional $\supset$ reflects the view that if the antecedent is not true, then we cannot tell the value of the conditional. However, $\rightarrow$ parallels the usual material conditional. The disjunction operator $\vee$ is the usual ‘max’.

The connectives $\neg$, $\rightarrow$ and $\vee$ are ‘pseudo-classical’ operators; if we map all the designated values to ‘true’ and the undesigned values to ‘false’, then the truth-conditions for these operators correspond to the classical connectives. The connectives $\sim$ and $\supset$ are not pseudo-classical.

Given our semantic account, we may inquire about an appropriate proof theory. Note that if we restrict our language to $\sim$ and $\supset$, then there will be no logical truths (formulas that are always designated); any formula composed with just these two connectives will take the undesigned value U when all atomic formulas take value U. Hence a proof theory for tautologies would not be appropriate were our language so restricted. However, an axiomatization of the consequence relation would still be of interest, and would include such sequents as $A, A \supset B \vdash B$.

Although often presented as restrictions of classical systems, many-valued logics can almost always be formulated as conservative extensions of classical logic. The trick is to include pseudo-classical connectives in the language.

A classical proof theory may be stated using the pseudo-classical connectives. Then using the autodescriptive operators, the additional required principles can be formulated. For example, some principles arise simply by stating the truth-conditions for each connective. The last entry on the table for $\supset$, above, corresponds to $J_F(A) \rightarrow (J_F(B) \rightarrow J_U(A \supset B))$ for an axiomatic system or to $J_F(A), J_F(B) \vdash J_U(A \supset B)$ for a sequent system.

3 Technical matters

All the metatheoretical concerns raised with classical logic also arise with respect to many-valued logics. The model theory of many-valued logics can be developed in much the same way as that of classical logic. Depending on the specifics of the syntax, proof theories are formulated in all the usual varieties: axiomatic, natural deduction, single- and multiple-conclusion consequence relations. Many-valued extensions of non-classical logics may be constructed, and many-valued versions of standard identity theory have been formulated.

Two frequently encountered questions warrant special mention, though we provide only partial answers here. (1) Given an arbitrary semantic theory, can we always find a corresponding proof theory? For the finite-valued case, the answer is affirmative, providing the syntax has appropriate connectives; but if the syntax is limited, it may be impossible to formulate an appropriate proof theory for a single-conclusion consequence relation. For the infinite-valued case, most interesting quantificational systems will not be axiomatizable; even for propositional logics it will be impossible to formulate a finitistic proof theory if the syntax is rich enough (for example, including autodescriptive operators). (2) Given a proof theory, can we always find a corresponding many-valued semantics? If we are concerned only with logical truth (not consequence in general) and if the proof theory admits a rule of substitution, then there will always be an infinite-valued semantics, though the designated values may turn out not to be decidable.

See also: [Logical and mathematical terms, glossary of](#); [Modal logic](#)

CHARLES G. MORGAN

References and further reading

- Belnap, N.D.** (1977) 'A Useful Four-Valued Logic', in J.M. Dunn and G. Epstein (eds) *Modern Uses of Multiple-Valued Logic*, Boston, MA: Reidel.(Uncertainty due to conflicting opinion.)
- Black, M.** (1937) 'Vagueness: An Exercise in Logical Analysis', *Philosophy of Science* 427-55.(Logic of vagueness.)
- Chang, C.C.** (1965) 'Infinite-Valued Logic as a Basis for Set Theory', in Y. Bar-Hillel (ed.) *Logic, Methodology and Philosophy of Science*, Amsterdam: North Holland.(Paradoxes of set theory.)
- Chang, C.C. and Keisler, H.J.** (1966) *Continuous Model Theory*, Princeton, NJ: Princeton University Press. (Many-valued model theory.)
- Dugundji, J.** (1940) 'Note on a Property of Matrices for Lewis and Langford's Calculi of Propositions', *Journal of Symbolic Logic* 5: 370-99.(Important modal systems do not have finite-valued semantics.)
- Fitting, M.C.** (1992) 'Many-Valued Modal Logics', parts 1 and 2, *Fundamenta Informaticae* 15: 235- 54, 17: 55-73.(Many-valued extensions of modal logics.)
- Fraassen, B.C. van** (1971) *Formal Semantics and Logic*, New York: Macmillan.(Finding a semantics for a given proof theory.)
- Kleene, S.C.** (1952) *Introduction to Metamathematics*, Princeton, NJ: Van Nostrand.(Uncertainty due to lack of termination of computation.)
- Martin, R.L.** (ed.) (1970) *The Paradox of the Liar*, New Haven, CT: Yale University Press.(On many-valued logics as a response to the liar paradox.)
- McCall, S.** (ed.) (1967) *Polish Logic 1920-1939*, London: Oxford University Press.(Source for papers of Łukasiewicz.)
- Morgan, C.G.** (1974) 'A Theory of Equality for a Class of Many-Valued Logics', *Zeitschrift für mathematische Logik* 20: 427-32.(Many-valued identity theory.)
- Morgan, C.G.** (1975) 'Similarity as a Theory of Graded Equality for a Class of Many-Valued Logics', in *Proceedings of the 1975 International Symposium of Multiple-Valued Logic*, Long Beach, CA: IEEE Society, 436-49. (Many-valued identity theory.)
- Morgan, C.G.** (1976) 'Many-Valued Propositional Intuitionism', in *Proceedings of the Sixth International Symposium on Multiple-Valued Logic*, Long Beach, CA: IEEE Society, 150-6.(Many-valued extension of

- intuitionistic logic.)
- Morgan, C.G.** (1979) 'Local and Global Operators and Many-Valued Modal Logics', *Notre Dame Journal of Formal Logic* 20: 401-11.(Many-valued extensions of common modal logics.)
- Morgan, C.G. and Pelletier, F.J.** (1976) 'Some Notes concerning Fuzzy Logics', *Linguistics and Philosophy* 1: 79-97.(Finding proof theory for infinite-valued semantics.)
- Putnam, H.** (1957) 'Three-Valued Logic', *Philosophical Studies* 8: 73-80.(On uncertainty due to the incompleteness of human knowledge.)
- Reichenbach, H.** (1944) *Philosophic Foundations of Quantum Mechanics*, Berkeley, CA: University of California Press.(On uncertainty due to statistical fluctuations in nature.)
- Rescher, N.** (1969) *Many-Valued Logic*, New York: McGraw-Hill.(Survey of historically important systems.)
- Rosser, J.B. and Turquette, A.R.** (1952) *Many-Valued Logics*, Amsterdam: North Holland.(Finding proof theory for finite-valued semantics.)
- Scarpellini, B.** (1962) 'Die Nicht-Axiomatisierbarkeit des unendlichwertigen Prädikatenkalküls von Łukasiewicz', *Journal of Symbolic Logic* 27: 159-70.(Finding proof theory for infinite-valued semantics.)
- Scott, D.** (1974) 'Completeness and Axiomatizability in Many-Valued Logic', in *Proceedings of the Tarski Symposium, Proceedings of Symposia in Pure Mathematics*, vol. 25, Providence, RI: American Mathematical Society, 411-35.(On uncertainty due to accumulating error.)
- Shoesmith, D.J. and Smiley, T.J.** (1978) *Multiple-Conclusion Logic*, Cambridge: Cambridge University Press.(Proof theory for a given semantics.)
- Thomason, S.K.** (1978) 'Possible Worlds and Many Truth Values', *Studia Logica* 37: 195-204.(Many-valued extensions of indexical logics.)
- Urquhart, A.** (1986) 'Many-Valued Logic', in D. Gabbay and F. Guenther (eds) *Handbook of Philosophical Logic*, vol. 3, Dordrecht: Reidel, 71-116.(Survey of important technical results.)

Many-valued logics, philosophical issues in

The first philosophically-motivated use of many-valued truth tables arose with Jan Łukasiewicz in the 1920s. What exercised Łukasiewicz was a worry that the principle of bivalence, ‘every statement is either true or false’, involves an undesirable commitment to fatalism. Should not statements about the future whose eventual truth or falsity depends on the actions of free agents be given some third status - ‘indeterminate’, say - as opposed to being (now) regarded as determinately true or determinately false? To implement this idea in the context of the language of sentential logic (with conjunction, disjunction, implication and negation), we need to show - if the usual style of treatment of such connectives in a bivalent setting is to be followed - how the status of a compound formula is determined by the status of its components.

Łukasiewicz’s decision as to how the appropriate three-valued truth-functions should look is recorded in truth tables in which (determinate) truth and falsity are represented by ‘1’ and ‘3’ respectively, with ‘2’ for indeterminacy (see tables in the main body of the entry). Consider the formula $A \vee B$ (‘A or B’), for example, when A has the value 2 and B has the value 1. The value of $A \vee B$ is 1, reasonably enough, since if A’s eventual truth or falsity depends on how people freely act, but B is determinately true already, then $A \vee B$ is already true independently of such free action. There are no constraints as to which values may be assigned to propositional variables. The law of excluded middle is invalidated in the case of indeterminacy: if p is assigned the value 2, then $p \vee \neg p$ also has the value 2. This reflects Łukasiewicz’s idea that such disjunctions as ‘Either I shall die in a plane crash on January 1, 2030 or I shall not die in a plane crash on January 1, 2030’ should not be counted as logical truths, on pain of incurring the fatalistic commitments already alluded to.

Together with the choice of designated elements (which play the role in determining validity played by truth in the bivalent setting), Łukasiewicz’s tables constitute a (logical) matrix. An alternative three-element matrix, the 1-Kleene matrix, involves putting $2 \rightarrow 2 = 2$, leaving everything else unchanged. And a third such matrix, the 1,2-Kleene matrix, differs from this in taking as designated the set of values $\{1, 2\}$ rather than $\{1\}$. The 1-Kleene matrix has been proposed for the semantics of vagueness. In the case of a sentence applying a vague predicate, such as ‘young’, to an individual, the idea is that if the individual is a borderline case of the predicate (not definitely young, and not definitely not young, to use our example) then the value 2 is appropriate, while 1 and 3 are reserved for definite truths and falsehoods, respectively. Łukasiewicz also explored, as a technical curiosity, n-valued tables constructed on the same model, for higher values of n, as well as certain infinitely many-valued tables. Variations on this theme have included acknowledging as many values as there are real numbers, with similar applications to vagueness and approximation in mind.

1 Second thoughts on some popular applications

Motivated by a concern that the principle of bivalence (‘every statement is either true or false’) involves an undesirable commitment to fatalism, Łukasiewicz introduced truth tables involving the value ‘indeterminate’ in addition to ‘true’ and ‘false’ in the 1920s (see Łukasiewicz 1970). His decision as to how these three truth-values should be combined is shown in the following tables:

$\wedge$	1	2	3	$\vee$	1	2	3
*1	1	2	3	1	1	1	1
2	2	2	3	2	1	2	2
3	3	3	3	3	1	2	3

$\rightarrow$	1	2	3	A	$\neg A$
1	1	2	3	1	3
2	1	1	2	2	2
3	1	1	1	3	1

The asterisk flagging the first occurrence of ‘1’ in the tables is the conventional way of indicating that this value is ‘designated’: in other words, that this value plays the role, in defining validity, played by truth in the bivalent setting.

The formal apparatus introduced by these tables is a special case of the notion of a matrix M with which are associated an algebra (a non-empty set - here, the set of matrix elements - closed under some specified operations), that we may call $A(M)$; and a set of matrix elements which are to count as the designated elements, which we will call $D(M)$. An alternative three-element matrix, the 1-Kleene matrix, differs from Łukasiewicz's only in having $2 \rightarrow 2 = 2$. A third such matrix, the 1,2-Kleene matrix, differs from this in taking as designated the set of values $\{1, 2\}$ rather than $\{1\}$.

One might expect the term 'matrix' to apply to the individual tables for the connectives depicted in such diagrams, but the standard logical usage, reflected in the definition just given, means that it is the whole set of tables that depicts a single matrix. If we call this matrix M , then the set of M -elements is $\{1, 2, 3\}$ and the algebra $A(M)$ has operations ' $\wedge$ ', ' $\vee$ ', ' $\rightarrow$ ', ' $\neg$ ', behaving as depicted in the tables. (When more care is called for, we distinguish these operations from the connectives, writing something like ' $\wedge^M$ ', ' $\vee^M$ ' and so on.) Finally, $D(M) = \{1\}$. A logic which is - to use terminology defined in §2 below - 'determined by' a matrix with n elements but by no matrix with fewer than n elements is called an n -valued logic.

From the table for ' $\vee$ ' we can see that $1 \vee 2 = 1$. While this seems satisfactory (if A is true then $A \vee B$ will be true regardless of the value assigned to B), we can see that all is not well by looking at another case. The table for ' $\wedge$ ' tells us that $2 \wedge 2 = 2$: a result which is suggested by the case in which the two conjuncts concerned are independent ('I shall die in a plane crash', 'Britain will become a republic'), as well as by the case in which the conjuncts are identical, but which goes obviously wrong otherwise ('I shall die in a plane crash', 'I shall not die in a plane crash' - to take an extreme case). The falsity of the conjunction 'I shall die in a plane crash and I shall not die in a plane crash' is fixed regardless of how anyone decides to act. Thus it should receive the value 3. But of course we cannot simply emend Łukasiewicz's matrix and insert a '3' in the middle box of the ' $\wedge$ ' table, since we shall then have the wrong result for the other cases. Similar reasoning applies in the case of disjunction: the disjunction of 'I shall die in a plane crash' and its negation is not something whose eventual truth can be prevented, so it should have the value 1. So we cannot acquiesce in Łukasiewicz's $2 \vee 2 = 2$. The only conclusion possible is that treating the threefold classification Łukasiewicz had in mind as a range of values with respect to which the familiarly truth-functional connectives could indeed be taken as truth-functional is misconceived. Not every threefold classification of statements gives us a three-valued logic.

It seems appropriate to return a similarly negative verdict on the proposed application of 3- (and more) element matrices, including the 1-Kleene matrix (see [Cleave 1974](#)), to vagueness (see [Vagueness §3](#)). In these and other cases the truth-functional paradigm is simply inappropriate for the notions being modelled ([Geach 1949](#)). The problem of failed truth-functionality which afflicts these approaches arises out of the fact that there is a linear ordering of the truth-values with respect to which ' $\wedge$ ' and ' $\vee$ ' are interpreted as *min* and *max*. In the case of Łukasiewicz's three-valued tables - contrary to what is suggested by the labelling - we are taking 1 as the greatest and 3 as the least, with 2 intermediate. Not all many-valued tables exhibit this kind of linear ordering, however, including some that Łukasiewicz himself proposed for treating modal notions; typically, though, there are more suggestive ways of developing the semantics for such applications than by the laying down of such tables (see [Prior 1955](#)). We return in §4 to other applications; we shall also have occasion to observe that the above talk of rejecting the principle of bivalence turns out on reflection to be somewhat dubious - even apropos of certain more promising applications of matrix semantics than those touched on above.

2 Logic

Suppose we have a matrix M and a formal language whose primitive connectives are in one-to-one correspondence with the operations of $A(M)$. Then an assignment of elements of M to the formulas of that language, in which $A(M)$ is used to compute the value of compounds on the basis of their components, will be called an M -evaluation. This generalizes the idea of a 'valuation', which term we shall always reserve for the bivalent case (so a valuation assigns either T or F to every formula), and we wish now to adapt the familiar idea of validity as preservation of truth. The resulting notion is to apply to sequents, and we begin by distinguishing three conceptions of what a sequent is to look like - three (as we shall say) 'logical frameworks'.

The most general framework we call SET-SET, and it takes a sequent of a given language to be something of the form $\Gamma \succ \Delta$ (read ' Γ , therefore Δ ' or ' Γ yields Δ '), where Γ and Δ are arbitrary sets of formulas of that language ([Blamey 1986](#)). If we impose the condition that Δ has to be $\{B\}$ for some formula B , we have the more

restricted logical framework SET-FMLA, and write $\Gamma \succ B$ rather than $\Gamma \succ \{B\}$. Finally, if we impose, in addition to this condition on Δ , the further condition that Γ must be empty, we have the framework we call FMLA, and when no confusion arises we identify the sequents $\succ B$ (that is, $\emptyset \succ B$) which this framework provides with the formulas B themselves.

Early approaches to logic (Frege, Hilbert) used the framework FMLA, syntactically codifying a logic with the axiomatic method. The provable formulas were formal analogues of the ‘logical truths’ of that area of natural language being formalized. The sequents of SET-FMLA are formal analogues of arguments with premises on the left of the ‘ $\succ$ ’, and the conclusion to the right. For SET-SET, we have a generalization in which, roughly speaking, we may say that sets of formulas on the right are understood disjunctively, just as those on the left are (as in SET-FMLA) understood conjunctively. More precisely, in the bivalent setting, we say that a sequent $\Gamma \succ \Delta$ of SET-SET holds on a valuation v (for the language in question) just in case if $v(A) = \text{T}$ for every $A \in \Gamma$, then $v(B) = \text{T}$ for some $B \in \Delta$. Since the other logical frameworks simply restrict the class of sequents, we may use this same definition in connection with them. A set of sequents which contains all substitution instances of any of its elements will be what we have in mind as a *logic* in that framework. The set of sequents determined by a class of valuations is the set comprising precisely those sequents holding on each valuation in the class. This set may or may not constitute a logic, on the definition just given, depending on what class of valuations is chosen.

It remains to provide a ‘matrix semantics’ version of the above. We simply make clear the intention that the designated values are those which play the role of truth. Suppose, then, that h is an M -evaluation, for some matrix M ; we say that the sequent $\Gamma \succ \Delta$ holds on h just in case

If $h(\Gamma) \subseteq D(M)$ then $h(\Delta) \cap D(M) \neq \emptyset$,

where $h(\Gamma)$ denotes $\{h(A) \mid A \in \Gamma\}$. Again, this definition fixes what it is for a sequent of SET-FMLA or FMLA to hold on an evaluation. Finally, a sequent is *valid* in a matrix M if it holds on every M -evaluation. We will not use the term ‘tautologous sequent’ (or ‘tautology’, in the case of FMLA) except for the special case of validity in the familiar two-element matrix. For this special case, also, we do not need the distinction between valuations and evaluations. The set of sequents determined by a class of matrices comprises precisely those sequents valid in each matrix in the class: such a set will always be a logic (as defined above). A logic determined by the class $\{M\}$ will be said to be determined by the matrix M . Notice that the logic in FMLA determined by the 1-Kleene matrix (which differs only from Łukasiewicz in that $2 \rightarrow 2 = 2$) is just the empty set, the assignment of the value 2 to every formula being an evaluation for this matrix. Since Łukasiewicz worked in FMLA, this is one reason that in his tables we have $2 \rightarrow 2 = 1$ (rather than $= 2$). (The logics determined by the Kleene matrices in the other frameworks are of course non-empty.)

3 Logic (cont.)

We say a logic is ‘many-valued in the narrow sense’ when it is determined by some matrix, and ‘many-valued in the broad sense’ when it is determined by some class of matrices. In FMLA this is a distinction without a difference, an early result of Lindenbaum and Tarski having established that in this framework every logic is many-valued in the narrow sense. However, in the case of the other frameworks, the distinction is a real one.

In SET-FMLA, for example, every logic closed under certain structural rules (‘overlap’, ‘dilution’ and ‘cut’ (for sets) - see Shoesmith and Smiley 1978) is broadly many-valued. We call a relation ‘ $\vdash$ ’ between sets of formulas and individual formulas (of some given language) a consequence relation (on that language) when there is some collection Σ of sequents of SET-FMLA which is closed under the SET-FMLA structural rules and for which we have $\Gamma \vdash A$ if and only if the sequent $\Gamma \succ A$ belongs to Σ . (This means that ‘ $\Gamma \succ A$ ’ is treated as an alternative notation for $\langle \Gamma, A \rangle$.) Closure under these structural rules is a necessary condition for a logic to be determined by any class of valuations. These same rules must be satisfied by any logic determined by a class of matrices for the following reason. Given an M -evaluation h (for some matrix M), we define the (bivalent) valuation v_h induced by h thus: $v_h(A) = \text{T}$ if and only if $h(A) \in D(M)$, for all formulas A . Then a sequent holds in the many-valued sense on h just in case it holds, in the bivalent sense, on v_h . Accordingly a set of sequents is determined (many-valued sense) by a class $\mathcal{C}$ of matrices precisely when it is determined (bivalent sense) by the class of valuations $\{v_h: h \text{ is an } M\text{-evaluation for some } M \in \mathcal{C}\}$. Thus the structural rules must also be satisfied by

any SET-FMLA logic which is many-valued in the broad sense. Now while a necessary and sufficient condition for a logic in SET-FMLA to be many-valued in the broad sense is that it should be a consequence relation, there is a further condition to be satisfied in the case of the narrow sense, which is called the cancellation condition in Shoesmith and Smiley (1978). (See also Wójcicki 1974.) This further condition is not satisfied in the case of the consequence relation determined by the pair comprising the 1-Kleene matrix and the 1,2-Kleene matrix, for example, which is not, therefore, a many-valued logic in the narrow sense.

Note that the narrow sense could be further narrowed to exclude determination by a two-element matrix: it is only on such a usage that Łukasiewicz could be counted a pioneer of many-valued logic, for example. A further narrowing in the ‘opposite’ direction comes by disallowing determination by an infinite matrix to suffice for many-valuedness. The motivation for excluding infinite matrices comes from more than one source. Philosophically, there may be a feeling that such a case involves too many distinctions for the values so distinguished to represent a plausible generalization of the idea of truth-values (Haack 1978: 214). Perhaps the assignment made by an evaluation in this case is better thought of as the assignment of a ‘proposition’ to a formula (see Dunn 1975). There is also a technical pressure tending towards excluding the infinite case, coming from the Lindenbaum-Tarski result mentioned above together with an early concentration on the framework FMLA: since every logic is many-valued in the absence of the restriction to finite determining matrices, we might impose this restriction to avoid making the term ‘many-valued’ vacuous. However, since (again) this threat of vacuity does not arise for other logical frameworks, we prefer not to narrow further the above narrow sense of ‘many-valued’. It is better to have a separate word for the property of being determined by a finite matrix; the usual word is ‘tabular’.

The first demonstration that a logic was not tabular came when Gödel (1932) addressed this question for intuitionistic sentential logic in FMLA. We can introduce the argument he gave for this conclusion by considering a well-known puzzle. Suppose you have a drawer containing (only) twenty socks, ten of which are blue and ten of which are green. You have to choose a pair of the same colour to wear, but because the room is dark, you cannot see the colours. The question is: what is the smallest number of socks you need to remove in order to be sure that among them there will be a matching pair? And the answer is: (not *eleven* but) *three*. The reason is that, since there are only two colours involved, three socks cannot all be of different colours, so you are guaranteed a matching pair. This is a special case of what is sometimes called the ‘pigeonhole principle’: if a number of pigeons are to be placed in a smaller number of pigeonholes, then at least one pigeonhole must have more than one pigeon in it.

With the example of the socks in mind, consider the question of whether or not the following formula is a (two-valued) tautology:

$$Pigeon_3 \quad (p \leftrightarrow q) \vee (p \leftrightarrow r) \vee (q \leftrightarrow r).$$

It should not be necessary to construct the eight-line truth table to answer this question. Since there are only two truth-values (T, F) to go round, no valuation can assign different truth-values to all three propositional variables p , q and r . (The variables here are playing the role of the pigeons - whence the labelling of the formula - with the truth-values being the pigeonholes.) So each valuation must assign the value T to at least one of the disjuncts of the above formula, thus assigning T to the whole disjunction, which is therefore a tautology.

We turn to Gödel’s argument that intuitionistic logic in FMLA - which we shall denote by IL - is not determined by any finite matrix. Let CL denote classical sentential logic in FMLA (alias the set of tautologies in two-valued logic). To begin with, let us see how considerations such as those of the preceding paragraph reveal IL to be determined by no three-element matrix, when supplemented by three additional facts: (1) $IL \subseteq CL$; (2) IL has the ‘disjunction property’: for all formulas A, B , if $A \vee B \in IL$, then either $A \in IL$ or $B \in IL$; and (3) if C is a disjunction one of whose disjuncts is a formula of the form $A \leftrightarrow A$, then $C \in IL$. (We take ‘ $\leftrightarrow$ ’ as a primitive connective for simplicity here; in fact the whole argument goes through if this is replaced by ‘ $\rightarrow$ ’.) Now consider the formula

$$Pigeon_4 \quad (p \leftrightarrow q) \vee (p \leftrightarrow r) \vee (p \leftrightarrow s) \vee (q \leftrightarrow r) \vee (q \leftrightarrow s) \vee (r \leftrightarrow s).$$

This formula is not provable in IL, in view of facts (1) and (2) above. Now suppose that some three-element matrix M determines IL. For any M -evaluation h , fact (3) and the pigeonhole principle imply that $Pigeon_4$ holds on h , since h cannot assign distinct M -elements to all four propositional variables. Thus, though not IL-provable, $Pigeon_4$ is M -valid: so IL is not determined by M . The same argument shows (by analogous consideration of the

formula we naturally call *Pigeon_n*) that, for any finite number n , IL is not determined by any matrix having n elements: IL is not tabular.

4 From evaluations to valuations

Our introduction, in §3, of the valuation v_h induced by a matrix evaluation h , serves as a reminder of the twofold division which is always present in the background of the many-valued setting, in the form of the ubiquitous designated/undesignated distinction. The notion of validity for sequents is always a matter of invariable possession or preservation of truth: of T as a value of the functions v_h . Distinctions within the classes of designated and undesignated matrix-values do not matter for validity, and enter the semantic account only because the connectives concerned cannot be seen as simply ‘designation-functional’: in the case of Łukasiewicz’s and Kleene’s three-valued tables, for example, while $h(\neg A)$ is uniquely determined by $h(A)$, there is no such unique determination of $v_h(\neg A)$ by $v_h(A)$. So we do our compositional semantics at the level of evaluations, and then modulate to valuations for talk of validity.

Here is an analogy: imagine that chemistry were concerned in the first instance with the colours of chemical compounds. We would like to know what colour a compound of substances of given colours will be. The trouble is that just knowing what colour the constituents are is not enough to tell us what colour the compound will be. We also need to know about their chemical composition, so that this extra information can be used to predict the eventual colour of the compound, via its chemical composition. Similarly, even if all we are interested in knowing about a composite formula $A \# B$ (where $\#$ is some binary connective) is whether or not it has a designated value on evaluation h , we cannot obtain this information - that is, the value of $v_h(A \# B)$ - just from knowing $v_h(A)$ and $v_h(B)$ (the ‘colours’ of A and B). Rather, we need to know $h(A)$ and $h(B)$, so that we can apply our matrix operation $\#$ to these values and obtain $h(A \# B)$, and then, eventually, $v_h(A \# B)$. This is precisely to acknowledge the possibility of a failure of what was called designation-functionality above.

But are the subdivisions effected by the apparatus of many-valued logic of merely instrumental interest? Such a suggestion has been elaborated over a long period by Michael Dummett, beginning with his article ‘Truth’ in 1959. The problem Dummett noted is that when we consider the point of the true/false dichotomy as applied to whole sentences (of an interpreted language such as English) which might be used to make assertions, there does not seem to be room for alleging any particular sentence to fall on neither side:

We need to distinguish those states of affairs such that if the speaker envisaged them as possibilities he would be held to be either misusing the statement or misleading his hearers, and those of which this is not the case: and one way of using the words ‘true’ and ‘false’ would be to call states of affairs of the former kind those in which the statement was false, and the others those in which the statement was true.

([1959] 1978: 11-12)

Dummett went on to describe in the light of this criterion how particular cases in which statements have been described as neither true nor false should be classified as true or as false, depending on the nature of the case. The various values of many-valued matrices are then seen as subdivisions of the true (the designated values) and the false (the undesignated) - different ways of being true or of being false, the difference between which needs to be taken into account for settling how the fundamental binary true/false distinction is to apply to compounds into which they enter as components. We repeat one of Dummett’s illustrations:

I once imagined a case in which a language contained a negation operator ‘ $\neg$ ’ which functioned much like our negation save that it made ‘ $\neg(A \rightarrow B)$ ’ equivalent to ‘ $A \rightarrow \neg B$ ’, where ‘ $\rightarrow$ ’ is the ordinary two-valued implication. In this case, the truth or falsity of ‘ $\neg(A \rightarrow B)$ ’ would not depend solely on the truth or falsity of ‘ $A \rightarrow B$ ’, but on the particular way in which ‘ $A \rightarrow B$ ’ was true (whether by the truth of both constituents or by the falsity of the antecedent). This would involve the use of three-valued truth tables, distinguishing two kinds of truth. In the same way, it might be necessary to distinguish two kinds of falsity.

([1963] 1978: 155)

Indeed, Dummett (1959) suggested this last possibility as the appropriate gloss on the response of those who describe sentences containing non-denoting singular terms as ‘neither true nor false’. The response is a natural one if one thinks of falsity of a sentence as the truth of its negation (supposing the identification of a sentence’s

negation is not itself problematic), but since it would be equally erroneous ('misleading') to assert either the given afflicted sentence or its negation, in terms of the more fundamental true/false dichotomy, both sentences fall alike on the false side. The two undesigned values of Łukasiewicz's three-valued logic can be regarded in the same light: as cases in which a sentence having either value are not, the facts being what they are, fit for assertion. They differ in respect of whether the negation of the sentence in question suffers the same fate (according to Łukasiewicz's metaphysical assumptions, that is, and setting to one side the problems about truth-functionality mentioned in §1).

It is for reasons such as this that Dummett speaks of 'the legitimate but comparatively superficial character of the rejection of the law of bivalence involved in the use of a many-valued logic' (1973: 448). In the same work, Dummett draws a distinction between the 'assertoric content' and the 'ingredient sense' of a statement, introduced in terms of the distinction between knowing the meaning of a statement 'in the sense of grasping the content of an assertion of it' and 'in the sense of knowing the contribution it makes to determining the content of a complex statement in which it is a constituent' (1973: 447). Sentences (or statements) which are alike in respect of assertoric content are true (that is, have some designated value) under the same conditions, which does not imply that the replacement of one by the other in some complex sentence will secure a similar likeness for the two complex sentences produced. When the complex sentences differ in respect of designation in this way, the components in respect of which they differ themselves differ in ingredient sense, their similarity of assertoric content notwithstanding. Thus, in terms of our formal notation, and slightly oversimplifying, equality of ingredient sense, for A and B , corresponds to having $h(A) = h(B)$ for every matrix-evaluation h , while equality of assertoric content corresponds to having $v_h(A) = v_h(B)$ for each such evaluation h .

Dummett's rationale for many-valued (that is, k -valued for $k > 2$) logic, then, is that there is sometimes a need to record differences in ingredient sense among statements which do not differ in assertoric content. Without going into too much technical detail, though, we can see that this is not quite accurate. We call a consequence relation ' $\vdash$ ' 'extensional' if for all formulas A, B , with $C(A)$ and $C(B)$ the results of putting formulas A and B into some context C , and for all sets Γ of formulas, we have

If $\Gamma, A \vdash B$ and $\Gamma, B \vdash A$, then $\Gamma, C(A) \vdash C(B)$;

and we call ' $\vdash$ ' 'congruential' when the above condition is satisfied by all A, B, C for the special case in which $\Gamma = \emptyset$. Congruentiality is strictly weaker than extensionality (consider normal modal logics, for example) and, as the end of the last paragraph makes clear, the need to distinguish ingredient sense from assertive content arises precisely when the logic (in SET-FMLA, for expository convenience) is not congruential. Yet, as the reader may care to check, it is failures of the stronger condition, extensionality, rather than just failures of congruentiality, that force a matrix treatment to involve more than two values.

5 From evaluations to valuations (cont.)

We end by briefly taking up one further issue raised by the above exposition: the case of intuitionistic logic. A perusal of Dummett (1959) reveals that his talk (quoted above) of the 'superficial character of the rejection of the law of bivalence' in many-valued logic is intended to contrast with the 'deeper' character of such a rejection in the case of intuitionistic logic. One reason for not considering intuitionistic logic to be many-valued seems hardly to the point here: namely the fact that, as we saw in §3, intuitionistic logic in FMLA is not determined by any finite matrix. (Nobody questions the classification of Łukasiewicz's infinite-valued logic as many-valued, for example.) By the Lindenbaum-Tarski result cited in that section, this ceases to be the case if 'finite' is deleted. Turning to SET-FMLA, the consequence relation associated with intuitionistic logic satisfies the cancellation condition (alluded to in §3) and so here too there is a single determining matrix. On formal grounds, then, we have to say that intuitionistic logic in either of these frameworks is indeed many-valued in the narrow sense of the term. Cardinality of determining matrices aside, are there any good reasons for not regarding IL as many-valued? There is a sense of 'many-valued' which we have not touched on, which allows an affirmative answer to this question: but it has more to do with what its adherents say about a logic than with the logic itself. The usual sort of reason for denying that IL is many-valued is that it would be inconsistent, according to IL itself, for an intuitionist to claim of some given statement that it was neither true nor false - at least when the falsity of A is identified with the truth of $\neg A$. (This use of 'falsity' is of course not that focused on in the quotation from Dummett (1959).) For

such a claim would then be of the form $\neg(A \vee \neg A)$, or, IL-equivalently, of the form $\neg A \wedge \neg\neg A$: but the negation of anything of this form is itself IL-provable. So the claim is not one an intuitionist could consistently make. The trouble with this as a way of distinguishing IL from a paradigmatically many-valued logic such as Łukasiewicz's three-valued logic is that dangerously similar moves can be made in the context of the latter logic also. If, instead of adopting an external perspective and saying that A is a statement which is neither true nor false, we think of representing such a claim as the claim that for the given A (which concerns, say, someone's later free actions), $\neg(A \vee \neg A)$, then again we have the equivalence with $\neg A \wedge \neg\neg A$, and this time the further equivalence with $\neg A \wedge A$. And while the negation of the latter is not provable in Łukasiewicz's logic (so that here we have a disanalogy with the case of IL) it remains the case that nothing of that form can receive the designated value on any matrix evaluation: so it is presumably after all not the kind of claim anyone attracted to Łukasiewicz's revision of classical logic ought to be making. As in the case of IL, such a theorist is distinguished from an adherent of classical logic by refraining from making certain claims (for example, some having the form $A \vee \neg A$), rather than by trying to make contrary claims. Further work, then, is called for on the part of anyone wishing to maintain that in some philosophically significant sense Łukasiewicz's logic is many-valued while that of the intuitionists is not.

See also: [Intuitionism](#); [Logical and mathematical terms, glossary of](#); [Łukasiewicz](#); [Many-valued logics](#)

LLOYD HUMBERSTONE

References and further reading

- Blamey, S.R.** (1986) 'Partial Logic', in D. Gabbay and F. Guenther (eds) *Handbook of Philosophical Logic*, vol. 3, *Alternatives to Classical Logic*, Dordrecht: Reidel, 1-70. (The logic in SET-SET determined by the 1-Kleene and 1,2-Kleene matrices taken together is the $\{\wedge, \vee, \neg, \rightarrow\}$ -fragment of the 'partial logic' presented here, although Blamey urges us to take the 'indeterminate' not as a truth-value on a par with the other two, but as representing undefinedness in respect of truth-value.)
- Cleave, J.P.** (1974) 'The Notion of Logical Consequence in the Logic of Inexact Predicates', *Zeitschrift für mathematische Logik und Grundlagen der Mathematik* 20: 307-24. (Proposes an application of the Kleene matrices to the logic of vagueness.)
- Czelakowski, J.** (1981) 'Equivalential Logics (I), (II)', *Studia Logica* 40: 227-36, 355-72. (Includes a discussion bearing on what we call 'extensionality' in §4.)
- Dummett, M.A.E.** (1959) 'Truth', *Proceedings of the Aristotelian Society* 59: 141-62; repr. with postscript (1972) in *Truth and Other Enigmas*, London: Duckworth, 1978. (Distinguishes deep from superficial grounds for rejecting the principle of bivalence.)
- Dummett, M.A.E.** (1963) 'Realism', in *Truth and Other Enigmas*, London: Duckworth, 1978, 145-65. (Further elaboration of the distinction between deep and superficial grounds for rejecting the principle of bivalence.)
- Dummett, M.A.E.** (1973) *Frege: Philosophy of Language*, London: Duckworth. (See the appendix to chapter 12 for a critical review of some central concepts of matrix semantics, using some traditional terminology not employed here.)
- Dummett, M.A.E.** (1991) *The Logical Basis of Metaphysics*, Cambridge, MA: Harvard University Press, esp. 48. (More on assertoric content versus ingredient sense.)
- Dunn, J.M.** (1975) 'Intensional Algebras', in A.R. Anderson and N.D. Belnap, *Entailment: The Logic of Relevance and Necessity*, vol. 1, Princeton, NJ: Princeton University Press, 1975, 180-206. (A gentle introduction to algebraic semantics.)
- Fine, K.** (1975) 'Vagueness, Truth, and Logic', *Synthese* 30: 265-300. (This paper, like Kamp's (1975), includes a good formulation of the 'failed truth-functionality' argument, as in §1 above, and an explanation of the appeal of a different treatment of vagueness, in terms of 'supervaluations'.)
- Geach, P.T.** (1949) 'If's and And's', *Analysis* 9: 58-62. (First published appearance of the 'failed truth-functionality' argument.)
- Gödel, K.** (1932) 'Zum intuitionistischen Aussagenkalkül', *Anzeiger der Akademie der Wissenschaften Wien, mathematisch-naturwissenschaftliche Klasse* 69: 65-6; trans. 'On the Intuitionistic Propositional Calculus', in *Kurt Gödel: Collected Works*, vol. 1, *Publications 1929-1936*, ed. S. Feferman, J.W. Dawson Jr, S.C. Kleene, G.H. Moore, R.M. Solovay and J. van Heijenoort, Oxford: Oxford University Press, 1986, 223-5. (In this paper Gödel shows that intuitionistic logic in FMLA is not tabular. A by-product of the argument is a proof of the existence of an infinite sequence of distinct logics intermediate between classical and intuitionistic propositional logic.)
- Haack, S.** (1978) *Philosophy of Logics*, Cambridge: Cambridge University Press. (Chapter 11 provides a short

survey of some intended applications of many-valued logic.)

- Kamp, H.** (1975) 'Two Theories about Adjectives', in E.L. Keenan (ed.) *Formal Semantics of Natural Language*, Cambridge: Cambridge University Press, 123-55. (This paper, like Fine's (1975), includes a good formulation of the 'failed truth-functionality' argument, as in §1 above, and an explanation of the appeal of a different treatment of vagueness, in terms of 'supervaluations'.)
- Łoś, J. and Suszko, R.** (1958) 'Remarks on Sentential Logics', *Indagationes Mathematicae* 20: 177-83. (This paper includes an ancestor of the cancellation condition of Shoesmith and Smiley (1978).)
- Łukasiewicz, J.** (1970) *Jan Łukasiewicz: Selected Works*, ed. L. Borkowski, Amsterdam: North Holland. (See especially 'On Three-Valued Logic' (1920), 'On Determinism' (Łukasiewicz's inaugural address at Warsaw University; 1922) and 'Philosophical Remarks on Many-Valued Systems of Propositional Logic' (1930).)
- Prior, A.N.** (1955) 'Many-Valued and Modal Systems: An Intuitive Approach', *Philosophical Review* 64: 626-30. (Shows how to relate certain many-valued matrices to the possible world's semantics for modal logic later popularized by Kripke.)
- Rescher, N.** (1969) *Many-Valued Logic*, New York: McGraw-Hill. (Provides extensive historical information on the subject, including more on the Kleene matrices, which are versions of Kleene's 'strong' three-valued truth tables, the 'weak' ones being associated with Bochvar.)
- Scott, D.S.** (1974) 'Completeness and Axiomatizability in Many-Valued Logic', in L. Henkin *et al.* (eds) *Proceedings of the Tarski Symposium*, Providence, RI: American Mathematical Society. (This work provides a useful introduction to semantics in terms of valuations, and proposes a treatment of many-valued logic in which the 'many values' are replaced by 'many valuations'; the logical framework is a finitized version of SET-SET.)
- Scott, D.S.** (1976) 'Does Many-Valued Logic Have Any Use?', in S. Körner (ed.) *Philosophy of Logic*, Oxford: Blackwell, 64-74. (A less technical discussion of the philosophical underpinnings of Scott 1974.)
- Seeskin, K.R.** (1971) 'Many-Valued Logic and Future Contingencies', *Logique et Analyse* 14: 759-73. (Includes the 'failed truth-functionality argument' mentioned in §1.)
- Seeskin, K.R.** (1974) 'Some Remarks on Truth and Bivalence', *Logique et Analyse* 17: 101-9. (Among other things, the author notes some features of Łukasiewicz's three-valued logic which are anomalous, given his informal motivation. These reflect what Blamey (1986) calls non-monotonicity (in the table for implication).)
- Shoesmith, D.J. and Smiley, T.J.** (1978) *Multiple-Conclusion Logic*, Cambridge: Cambridge University Press. (Includes a general discussion of many-valued logic and the theory of consequence relations. The cancellation conditions for SET-FMLA and SET-SET appear in chapter 15.)
- Smiley, T.J.** (1976) Comment on Scott (1976), in S. Körner (ed.) *Philosophy of Logic*, Oxford: Blackwell, 74-88. (This deals with the 'failed truth-functionality' point from §1.)
- Urquhart, A.** (1986) 'Many-Valued Logic', in D. Gabbay and F. Guenther (eds) *Handbook of Philosophical Logic*, vol. 3, *Alternatives to Classical Logic*, Dordrecht: Reidel, 71-116. (Good for information about certain aspects of many-valued logic, such as fundamental completeness, but makes the mistake of claiming that there is no distinction between - as we would put it - the broad and narrow senses of 'many-valued' in the framework SET-SET. See Wroński 1987 for correction.)
- Wójcicki, R.** (1974) 'Note on Deducibility and Many-Valuedness', *Journal of Symbolic Logic* 39: 563-6. (Traces the history of the Shoesmith-Smiley cancellation principle.)
- Wójcicki, R.** (1988) *Theory of Logical Calculi*, Dordrecht: Kluwer. (A comprehensive survey of matrix methodology. Congruentiality, mentioned in §4, is treated under the name 'self-extensionality'.)
- Wroński, A.** (1987) 'Remarks on a Survey Article on Many Valued Logic by A. Urquhart', *Studia Logica* 46: 275-8. (Corrects a serious error in Urquhart 1986.)
- Wroński, A.** (1987) 'Remarks on a Survey Article on Many Valued Logic by A. Urquhart', *Studia Logica* 46: 275-8. (Corrects a serious error in Urquhart (1986).)

Mass terms

Mass terms are words and phrases such as 'water', 'wood' and 'white wallpaper'. They are contrasted with count terms such as 'woman', 'word' and 'wild wildebeest'. Intuitively, mass terms refer to 'stuff'; count terms refer to 'objects'. Mass terms allow for measurement ('three kilos of wood', 'much water'); count terms allow for counting, quantifying and individuating ('three women', 'each word', 'that wildebeest over there').

Philosophical problems associated with mass terms include (1) distinguishing mass from count terms, (2) describing the semantics of sentences employing mass terms, and (3) explicating the ontology presupposed by our use of mass versus count terms. Associated with these philosophical issues - especially the third - are the meta-philosophical issues concerning the extent to which any investigation into the linguistic practices of speakers of a language can be used as evidence for how those speakers view 'reality'.

1 Distinguishing mass and count terms

The distinction between mass terms ('water', 'wood' and so on) and count terms ('woman', 'word' and so on) can be seen as syntactic, semantic or pragmatic. If the distinction is seen as syntactic, one might remark that mass terms occur with the quantifiers 'much' and 'little' and with the unstressed article 'some', that they are susceptible to measurement phrases such as 'litres of' and 'amount of', and that they do not exhibit a singular/plural distinction. Conversely, count expressions occur with the quantifiers 'each', 'every', 'many', 'several', 'few' and the stressed 'some', use the indefinite article 'a(n)', are susceptible to counting phrases such as 'five' and 'a score of', and exhibit a singular/plural dichotomy manifested in the count term itself and in agreement with the verb phrase.

If it is seen as semantic - a distinction between the different ways that mass and count terms refer - then one might remark that count expressions refer to discrete, well-delineated objects while mass terms refer without making it explicit how the referent is individuated (some have said that the referents of mass terms are continuous rather than discrete). This feature of mass reference gives rise to the 'cumulative reference test' (any sum of parts which are *M* is also *M*) and to the 'distributive [homogeneous] reference test' (any part of something which is *M* is also *M*).

If the mass/count distinction is seen as a pragmatic one, then one will look to how people use count terms to 'individuate' the world. This gives rise to such tests as whether there is a definite answer to the question 'How many *X*'s are there in such-and-such place?'. In the philosophical literature (following [Strawson 1959](#)), terms which pass this counting test are often called 'sortal terms' - although they are equally often called count terms - and ones that fail the test are called mass terms. In this literature, with its emphasis on the pragmatic notions of 'identifying' and 'individuating', it is common to deny that such terms as 'thing', 'object' and so on are sortal (count), there being no definite answer to the question 'How many things are in the room?'. This is so despite the fact that such terms clearly satisfy the syntactic criteria.

All these tests - the syntactic, the semantic and the pragmatic ones - have been challenged. Writers have pointed out that mass terms such as 'wood' can also be used as count: 'a wood' might designate oak or spruce, for example. And 'wildebeest' can be used as a mass term: 'He's not really a vegetarian; he eats wildebeest'. Furthermore, a universal grinder would take an object that an alleged count term referred to - a chair, for example - and grind it up into a powder so that then there would be chair all over the floor. (This last sentence uses 'chair' in a mass manner, thereby showing that the language already has this usage of any count term in the background. And this usage exists despite the fact that we could also have said 'There is wood all over the floor'.)

The criteria just mentioned are usually seen as applying to entire noun phrases as well as to the simple nouns themselves. Furthermore, some writers attempt to apply the mass/count distinction to other syntactic categories, especially adjectives, verbs and adverbs. The application to verbs is especially interesting (see [Mourelatos 1978](#)).

2 Semantics and ontology of mass terms

The problem with giving a formal semantic analysis of mass terms arises because first-order predicate logic appears to assume that the entities in the domain of quantification are individuals, so it only makes sense to characterize them with count nouns. When we say, in the quantifier idiom, 'For all *x*, if *x* is *F*...', it is apparently

assumed that the items in the domain have already been individuated. For if *F* were to be interpreted as 'snow', for example, what would be the values of *x*?

Famously, Quine (1960) held that mass terms are ambiguous: when in 'subject position' they are singular terms (names), but when in 'predicate position' they are general terms (predicates) which are 'true of each portion of the stuff in question, excluding only the parts too small to count'. As a name (when in subject position), Quine holds that a mass term 'differs none from such singular terms as "mama" ..., unless the scattered stuff that it names be denied the status of a single sprawling object'.

This proposal has not satisfied various authors, who have objected to the nonuniform treatment and to various logical consequences of this approach. For example, on Quine's analysis, 'Water is wet, and this puddle is water' does not imply 'This puddle is wet'; and 'Water is water' does not come out a logical truth. Writers after Quine have proposed many different approaches. Possibly the most popular alternatives involve mereology, according to which the main operator is 'is a part of'. Mass (and other) terms are taken to designate 'mereological wholes'. Some authors have grafted onto pure mereology a notion of 'having certain structural properties', so as to avoid the minimal parts problem alluded to in Quine. (The atoms, *inter alia*, which are part of water are 'too small' to count as water.) But these theories also have not satisfied all those involved in this area, usually because the treatment of certain logical inferences is thought incorrect: the formal semantic analyses do not mirror intuitive beliefs concerning logical consequence.

An alternative is to retain the idea that mass terms name some kind of object - a 'substance' - and to invoke a relational predicate such as 'is constituted of'. This presents a number of tricky issues and there are types of sentences for which such an analysis is not obviously suitable, but still various authors have adopted it. Besides the formal differences entailed by these two approaches (mereological calculus of individuals versus classical logic with a relational constant of constitution), there is an ontological difference, for mereological wholes are generally taken to be physical whereas substances or kinds are often viewed as abstract entities (see [Substance](#)).

Another formal semantics of mass terms invokes sets as their denotation. Differences among theorists can then be seen as differences about what the sets contain. One question on which theorists differ is whether the sets contain only 'minimal entities' - the smallest items to which the mass term refers (flakes, maybe, for 'snow'; the items and size vary according to the mass term in question) - or whether it should contain 'ordinary objects' (flakes, drifts, snowmen, snowballs and so on; any object which can be said to be snow). The former proposal has not gained many adherents due to the difficulty of specifying a set of 'minimal entities' for such mass terms as 'garbage', 'speed' and 'information'. The latter proposal runs into difficulties in trying to account for the denotation of definite noun phrases (NPs) such as 'the snow on the table'. It is generally not true that there is exactly one snow-thing on the table. (There is one ball and also many flakes making it up, for example.) So the only reasonable proposal is for the NP to designate *all* the snow-things on the table. But then certain measurement sentences - for example, 'The snow on the table weighs one kilo' - come out wrong, since we will count the same snow-entities many times over.

Theories of mass terms show a fundamental division between those that are committed to abstract substances and those that are physicalistic in nature, invoking mereological wholes. On the physicalistic side are those theories which propose that the ontologically basic objects are the minimal entities, those which claim that the larger entities are 'constructions' out of these minimal entities, and those theories which propose that all these entities are equally basic. On the other side of the gulf are the various styles of substance theories, which usually invoke a lattice structure of kinds. Such ontological issues are discussed in Pelletier and Schubert (1989) and in Burkhardt and Smith (1991).

See also: [Logical and mathematical terms, glossary of](#); [Mereology](#)

JEFFRY PELLETIER

References and further reading

Bunt, H. (1985) *Mass Terms and Model-Theoretic Semantics*, Cambridge: Cambridge University Press. (This is the longest and possibly most thorough work on mass terms. It is aimed at a computer implementation of a natural language understanding system which will include mass terms. It invokes 'ensemble theory', which is a type of atomic mereology.)

- Burkhardt, H. and Smith, B.** (1991) 'Mass Terms', in *Handbook of Metaphysics and Ontology*, Munich: Philosophia. (This entry goes into the metaphysical and ontological presuppositions of mass terms in more detail than is possible here. It includes some speculation about how the mass/count distinction might have arisen in natural language.)
- Lønning, J.-T.** (1987) 'Mass Terms and Quantification', *Linguistics and Philosophy* 10: 1-52. (An example of the direction that research into the mereological interpretation of mass terms has taken.)
- Mourelatos, A.** (1978) 'Events, Processes, and States', *Linguistics and Philosophy* 2: 415-34. (Mourelatos draws explicit attention to the relationships that hold between certain types of verbs and verb phrases and the mass/count distinction.)
- Pelletier, F.J. (ed.)** (1979) *Mass Terms: Some Philosophical Problems*, Dordrecht: Reidel. (This anthology includes many of the classic articles on the topic of mass terms, especially in the philosophical tradition. It also includes a comprehensive bibliography up to 1978.)
- Pelletier, F.J. and Schubert, L.K.** (1989) 'Mass Expressions', in D. Gabbay and F. Guentner (eds) *Handbook of Philosophical Logic*, Dordrecht: Reidel, vol. 4, 327-407, esp. 391-4. (The most thorough summary of the work done on mass terms, both linguistic and philosophical. It also proposes two theories to account for the phenomena, and includes a comprehensive bibliography of work from 1978 to 1988, plus earlier work where relevant.)
- Quine, W.V.** (1960) *Word and Object*, Cambridge, MA: MIT Press, 91-100. (Although this book has only a small section on mass terms, it formed the starting point of most future formal work on the topic. Quine proposed his 'dual analysis' here and asserted that mass terms come from an 'early, primitive stage in linguistic evolution'.)
- Quirk, R., Greenbaum, S., Leech, G. and Svartvik, J.** (1985) *A Comprehensive Grammar of Contemporary English*, London: Longman. (The most thorough descriptive account of the grammar of the English language, with a section about the mass/count distinction, especially as it is used in normal speech. It does not deal particularly with the formal semantics of the distinction.)
- Strawson, P.F.** (1959) *Individuals: An Essay in Descriptive Metaphysics*, London: Methuen. (It is here that the distinction between 'feature-placing universal' and 'sortal universal' was drawn. Many writers wished to put mass terms in the former category, giving rise in the philosophical literature to the opposition 'mass/sortal'. Other writers thought of mass terms as a type of sortal term, while still others argued for a third category.)

Mereology

Mereology is the theory of the part-whole relation and of derived operations such as the mereological sum. (The sum of several things is the smallest thing of which they are all parts.) It was introduced by Leśniewski to avoid Russell's paradox.

Unlike the set-membership relation, the part-whole relation is transitive. This makes mereology much weaker than set theory, but gives the advantage of ontological parsimony. For example, mereology does not posit the proliferation of entities found in set theory, such as $\emptyset, \{\emptyset\}, \{\{\emptyset\}\}, \dots$

Mereology has occasioned controversy: over whether many things really have a mereological sum if they are either scattered or, even worse, of different categories; over the uniqueness of sums; and over Lewis' claim that the non-empty subsets of a set are literally parts of it.

1 Classical mereology

Mereology is the theory of the part-whole relation. It was introduced by Leśniewski (1916) to avoid Russell's paradox (see **Paradoxes of set and property**). Taking the part-whole relation as primitive, we define a proper part as a part which is not identical to the whole. We define disjointness to mean having no part in common. Then, in addition to the transitivity of the part-whole relation, the following axiom schema, unique composition, characterizes classical mereology.

If there is some F , there is precisely one thing (the sum or fusion of the F s) of which every F is a part and which has no part disjoint from every F .

The supplementation principle now follows, namely that if a is a proper part of b , then b has some proper part disjoint from a . For otherwise both a and b are sums of all the parts of a , contradicting uniqueness. It also follows that there is a sum of everything, U ; that any thing a which is not U has a complement, namely the sum of everything disjoint from a ; and that any two overlapping things a and b have a product (or intersection), namely the sum of all things which are parts of both a and b . Apart from the absence of a null individual, mereology has, with these operations, the structure of a Boolean algebra (see **Boolean algebra §1**).

Strictly speaking, what we take as the characteristic axiom schema will depend on the logic mereology is an extension of. In the above this is the predicate calculus with identity. Leonard and Goodman's formulation (the calculus of individuals) is an extension of (a fragment of) set theory. In that case, unique composition is the axiom that any non-empty set of individuals has a sum. Or suppose, like Lewis (1991), you are extending a logic which permits plural quantification. Then unique composition could be formulated as:

For any things there is one thing which has those things as parts and which has no part disjoint from those things.

Yet again, in a finitist mereology the formulation would be:

For any a and b there is a unique c of which a and b are parts and which has no part disjoint from a and b .

Classical mereology is, therefore, a family of theories, with unique composition adapted to the context.

Both the transitivity of the part-whole relation and unique composition have been rejected as counter-intuitive. Cruse (1979) gives an example, later discussed by Simons (1987: 107), which illustrates the supposedly counter-intuitive character of the transitivity of the part-whole relation: a handle is part of a door, a door is part of a house but, it might seem, the handle is not part of the house. This objection is based on a controversial interpretation of 'part' to mean 'significant part'.

Unique composition may be conveniently divided into the existence of sums and the uniqueness of these sums. Let us first consider uniqueness. The problem, as stated by Simons (1987:115-17), has to do with 'continuants', namely things which endure without having temporal parts. Continuants are themselves controversial, but let us suppose that cats are continuants. Now consider a whole cat, Tibbles, her tail, and the remainder of Tibbles, called Tib. If Tibbles loses her tail then, Simons argues, Tibbles and Tib will be non-identical things with precisely the

same parts. They are non-identical because they are identical, respectively, to Tibbles and Tib prior to tail loss, which are obviously non-identical (in other words, Tibbles is still Tibbles with or without her tail). A similar problem can be stated in modal rather than temporal terms. While there may be several solutions, we might well restrict classical mereology to categories of items which are not continuants. In that case Simons' nonclassical mereology would be appropriate for continuants.

The existence of sums of disparate collections has seemed too strong to many. Here it is important to note that mereology is demonstrably consistent, so there is no analogue of Russell's paradox. The objection, therefore, is not formal. The worst case is the existence of sums of items from different categories, such as: a kettle, the process in the kettle in which water boils, the spatiotemporal region in which the process occurs, the property of being at 100°C, and the relation of contiguity. Even though all those items might be thought of in connection with the activity of boiling water, it seems queer to say there is something which is the sum of them.

Lewis has defended the existence of disparate sums on the grounds that describing the sum is doing nothing more than describing the parts, for there is no extra ingredient which somehow unifies the parts into a whole (1991: 79-81). He also points out that attempts to avoid the sum of a scattered collection of items of the same category will fail because in such cases queerness is a matter of degree. This leaves open, however, the possibility of restricting unique composition to sums of items all taken from the same category.

2 Mereology, set theory and atoms

An atom may be defined as something which has no *proper* part. We may ask whether there are any atoms and whether every thing is a sum of atoms. If the answer to the first question is negative then a mereology is called 'atomless'; if the answer to the second is positive then it is 'atomistic'. Now mereology does not have to be either atomistic or atomless - it might, for instance, be the case that every object has proper parts even though the space they occupy is a sum of points. It seems rather less plausible that any one category of thing (for example, physical objects) should have both atoms and atomless parts. Mereology restricted to a single category would seem, therefore, to be either atomistic or atomless. For example, mereology restricted to spatiotemporal regions is atomistic if all regions are sums of points, but atomless if points are idealized limiting cases of arbitrarily small regions (see Whitehead 1919; Whitehead §2; see also Mortensen and Nerlich 1978; Clarke 1985).

Mereology is of theoretical interest partly because it has been used to provide either a replacement for, or a formulation of, the theory of sets. Because a sum of atoms has a unique decomposition into atoms, sums behave like sets of atoms. The chief difference between mereology and set theory is that two different sets of non-atoms, say, $\{a + b, c + d\}$ and $\{a + c, b + d\}$, can have the same sum, $a + b + c + d$. It was hoped that the resources of mereology would none the less be enough to replace sets in the foundations of mathematics. Thus Goodman and Quine (Goodman 1972: 173-98) showed how the ancestral (that is, a chain of instances) of a relation could be defined using the part-whole relation, avoiding set theory. They also showed how various statements such as 'There are exactly one-third as many Canadians as Mexicans' can be handled using mereology. Their approach ran into difficulties, however, with quantified numerical formulas, such as 'For all n , $n + n = 2n$ '.

Lewis, in his version of nominalistic set theory, adopted a rather different strategy, namely to imitate set theory using the resources of mereology (see Lewis 1991: 21-8). The chief difficulty seems to have been that without the resources of set theory the materials are lacking with which to construct enough pseudo-sets. Thus a sound Ockhamist reason for preferring mereology to set theory - its inability to generate infinitely many things out of just one thing - restricts its use as a substitute.

Rather different is Lewis' later programme of treating the theory of sets and proper classes as itself an extension of mereology (1991). Excepting the special case of the empty set, we may treat subclasses as parts and thus treat set theory as the extension of mereology obtained by adjoining the singleton operator. This operator forms unit sets, which are always atoms, out of individuals. Lewis is in favour of treating it as primitive, but, relying on results by Burgess and Hazen, he also shows how we can develop an alternative, structuralist account, according to which there are many possible singleton operators, between which set theory does not choose.

See also: Logical and mathematical terms, glossary of; Structuralism

PETER FORREST

References and further reading

- Burkhardt, H.** (1990) 'Mereology in Leibniz's Logic and Philosophy', *Topoi* 9: 3-13.(Historical background.)
- Clarke, B.L.** (1985) 'Individuals and Points', *Notre Dame Journal of Formal Logic* 26: 61-75. (Recent discussion of the mereology of space.)
- Cruse, D.A.** (1979) 'On the Transitivity of the Part-Whole Relation', *Journal of Linguistics* 15: 29-38.(An examination of the linguistic evidence for the failure of transitivity. The author makes a useful distinction between parts and attachments.)
- Eberle, R.A.** (1970) *Nominalistic Systems*, Dordrecht: Reidel.(Includes simplifications to mereology based on the assumption of atomism.)
- Goodman, N.** (1956) 'A World of Individuals', in I. Bocheński, A. Church and N. Goodman, *The Problem of Universals*, Notre Dame, IN: University of Notre Dame Press; repr. in *Problems and Projects*, Indianapolis, IN: Bobbs-Merrill, 1972.(Along with the work of Leśniewski and Whitehead, this is one of the classics of mereology.)
- Goodman, N.** (1972) *Problems and Projects*, Indianapolis, IN: Bobbs-Merrill.(§4 includes two important and accessible papers: 'A World of Individuals' and, with Quine, 'Steps Toward a Constructive Nominalism'.)
- Henry, D.P.** (1991) *Medieval Mereology*, Amsterdam: Benjamin.(Historical background.)
- Le Blanc, A.** (1985) 'New Axioms for Mereology', *Notre Dame Journal of Formal Logic* 26: 437-43. (One of a series discussing the axiomatics of mereology.)
- Lejewski, C.** (1983) 'A Note on Leśniewski's Axiom System for the Mereological Notion of Ingredient or Element', *Topoi* 2: 63-72.(A recent paper by a distinguished mereologist in the school of Leśniewski.)
- Leśniewski, S.** (1916) 'Podstawy ogólnej teorii mnogości I', *Prace Polskiego Koa Naukowe w Moskwie, Sekcja matematyczno-przyrodnicza* 2; trans. D.I. Barnett, 'Foundations of the General Theory of Sets I', in *Collected Works*, ed. S.J. Surma, J.T. Szrednicki and D.I. Barnett, Dordrecht: Kluwer, 1992, vol. 1, 129-73.(Classic paper on mereology. The word 'Sets' in the translation of the title is misleading. The paper is not about sets as we now understand them. Consult P. Simons' 'Discovering Leśniewski's Collected Works', *History of Philosophical Logic* 15: 227-35 (1994) for a useful list of typographical errors.)
- Lewis, D.K.** (1991) *Parts of Classes*, Oxford: Blackwell.(A lucid presentation of set theory as an extension of mereology, using the minimum of technicality; also includes an excellent introduction to mereology (based on plural quantification), a defence of classical mereology, a discussion of nominalistic set theory and some important, though technical, appendices with J.P. Burgess and A.P. Hazen.)
- Mortensen, C. and Nerlich, G.** (1978) 'Physical Topology', *Journal of Philosophical Logic* 7: 209-23. (Topology based on atomless mereology for space.)
- Simons, P.** (1987) *Parts: A Study in Ontology*, Oxford: Clarendon Press.(A most comprehensive work discussing both the Leśniewski and the Leonard and Goodman formulations of mereology, providing some criticisms of classical mereology and developing a nonclassical theory. Of necessity, much of the discussion is technical. Includes an extensive bibliography.)
- Smith, B.** (ed.) (1982) *Parts and Moments: Studies in Logic and Formal Ontology*, Munich: Philosophia.(Includes an extensive bibliography.)
- Smith, B.** (1985) *Parts and Wholes*, vol. 3, Stockholm: Forskningsrådsnämnden.(Includes a further bibliography on mereology.)
- Sobociński, B.** (1934) 'O kolejnych uproszczeniach aksjomatyki "ontologii" prof. S. Leśniewskiego', in *Fragmenty Filozoficzne. Księga Pamiątkowa ku Uczczeniu Piętnastolecia Pracy Nauczycielskiej w Uniwersytecie Warszawskim Profesora Tadeusza Kotarbińskiego*, Warsaw: Nakładem Uczniów, 144-60; trans. 'Successive Simplifications of the Axiom-System of Leśniewski's Ontology', in S. McCall (ed.) *Polish Logic 1920-1939*, Oxford: Clarendon Press, 1967.(One of a series, mostly in the *Notre Dame Journal of Formal Logic*, by mereologists in the school of Leśniewski, in which alternative axiomatics for mereology are discussed. For a survey, with references, see [Simons \(1987: ch. 2\)](#).)
- Szrednicki, J.T.J., Rickey, V.F. and Czelakowski, J. (eds)** (1984) *Leśniewski's Systems: Ontology and Mereology*, The Hague: Nijhoff.(Reprinted articles by mereologists in the tradition of Leśniewski, including Lejewski, Supecki and Sobociński.)
- Whitehead, A.N.** (1919) *An Enquiry concerning the Principles of Natural Knowledge*, Cambridge: Cambridge University Press.(Includes a formulation of mereology independent of, and almost as early as, Leśniewski's, with application to the theory of atomless spacetime.)

Modal logic

Modal logic, narrowly conceived, is the study of principles of reasoning involving necessity and possibility. More broadly, it encompasses a number of structurally similar inferential systems. In this sense, deontic logic (which concerns obligation, permission and related notions) and epistemic logic (which concerns knowledge and related notions) are branches of modal logic. Still more broadly, modal logic is the study of the class of all possible formal systems of this nature.

It is customary to take the language of modal logic to be that obtained by adding one-place operators ‘ $\Box$ ’ for necessity and ‘ $\Diamond$ ’ for possibility to the language of classical propositional or predicate logic. Necessity and possibility are interdefinable in the presence of negation:

$$\Box A \leftrightarrow \neg \Diamond \neg A \text{ and}$$

$$\Diamond A \leftrightarrow \neg \Box \neg A$$

hold. A modal logic is a set of formulas of this language that contains these biconditionals and meets three additional conditions: it contains all instances of theorems of classical logic; it is closed under modus ponens (that is, if it contains A and $A \rightarrow B$ it also contains B); and it is closed under substitution (that is, if it contains A then it contains any substitution instance of A ; any result of uniformly substituting formulas for sentence letters in A). To obtain a logic that adequately characterizes metaphysical necessity and possibility requires certain additional axiom and rule schemas:

$$K \quad \Box(A \rightarrow B) \rightarrow (\Box A \rightarrow \Box B)$$

$$T \quad \Box A \rightarrow A$$

$$5 \quad \Diamond A \rightarrow \Box \Diamond A$$

$$\text{Necessitation} \quad A / \Box A.$$

By adding these and one of the $\Box$ - $\Diamond$ biconditionals to a standard axiomatization of classical propositional logic one obtains an axiomatization of the most important modal logic, S5, so named because it is the logic generated by the fifth of the systems in Lewis and Langford’s *Symbolic Logic* (1932). S5 can be characterized more directly by possible-worlds models. Each such model specifies a set of possible worlds and assigns truth-values to atomic sentences relative to these worlds. Truth-values of classical compounds at a world w depend in the usual way on truth-values of their components. $\Box A$ is true at w if A is true at all worlds of the model; $\Diamond A$, if A is true at some world of the model. S5 comprises the formulas true at all worlds in all such models. Many modal logics weaker than S5 can be characterized by models which specify, besides a set of possible worlds, a relation of ‘accessibility’ or relative possibility on this set. $\Box A$ is true at a world w if A is true at all worlds accessible from w , that is, at all worlds that would be possible if w were actual. Of the schemas listed above, only K is true in all these models, but each of the others is true when accessibility meets an appropriate constraint.

The addition of modal operators to predicate logic poses additional conceptual and mathematical difficulties. On one conception a model for quantified modal logic specifies, besides a set of worlds, the set D_w of individuals that exist in w , for each world w . For example, $\exists x \Box A$ is true at w if there is some element of D_w that satisfies A in every possible world. If A is satisfied only by existent individuals in any given world $\exists x \Box A$ thus implies that there are necessary individuals; individuals that exist in every accessible possible world. If A is satisfied by non-existents there can be models and assignments that satisfy A , but not $\exists x A$. Consequently, on this conception modal predicate logic is not an extension of its classical counterpart.

The modern development of modal logic has been criticized on several grounds, and some philosophers have expressed scepticism about the intelligibility of the notion of necessity that it is supposed to describe.

1 History

Modal logic in the narrow sense was a topic of considerable interest to ancient and medieval philosophers. It occupied two chapters of Aristotle’s *De Interpretatione*, and a substantial part of the *Prior Analytics*. Discussion of argument forms involving necessity and possibility that included, and sometimes transcended, commentary on

Aristotle's was standard fare in Hellenistic and medieval treatises on logic (see [Logic, medieval](#)). From our vantage point the ancient and medieval discussion can be interpreted as including distinctions among various kinds of possibility and necessity and investigations of the logical relations among them, as well as logical investigations of the interactions between modalities and negation, modalities and conditionals or consequence, and modalities and quantifier expressions. Aristotle determines in *De Interpretatione*, for example, that 'It may be' and 'It cannot be' are contradictories, as are 'It may not be' and 'It cannot not be' (see [Logic, ancient](#)). Furthermore, 'from the proposition "It may be" it follows that it is not impossible' and in one sense 'the proposition "It may be" follows from the proposition "It is necessary that it should be"'. In another sense (which we might gloss as 'It is merely possible that'), 'It may be' is logically incompatible with 'It is necessary that it should be'.

Besides these purely modal principles, Aristotle and his commentators were concerned with arguments that we might think of as mixing time and modality. A notorious example is the fallacious 'sea battle' argument for determinism that he tries to debunk in *De Interpretatione* (see [Aristotle §20](#)). In addition to the admixture of temporal considerations, one should observe that the notion of necessity involved in these discussions is not likely to be the same as the one whose logical behaviour was summarized above. Aristotle himself catalogues four senses of the word 'necessary' in *Metaphysics* (V 5), and makes other distinctions elsewhere.

Although necessity and possibility have never ceased to play an important role in philosophical discourse, their logical properties were largely neglected in modern philosophy until the beginning of the twentieth century. The contemporary revival was sparked by C.I. Lewis' critique of Whitehead and Russell's *Principia Mathematica*. The logical system elaborated in *Principia* contained as theorems the formulas $p \supset (q \supset p)$ and $\sim p \supset (p \supset q)$, which Whitehead and Russell understood as asserting the apparently paradoxical propositions that if a sentence is true it is implied by any sentence and if a sentence is false it implies any sentence. Lewis maintained that these propositions, while unavoidable and unobjectionable with respect to Russell and Whitehead's understanding of implication, were false with respect to a more natural 'strict' sense of implication (see [Relevance logic and entailment §§1-2](#)).

Lewis embarked on a project of determining the appropriate axioms of strict implication with which to supplement this system. In *Principia* the 'material' implication $p \supset q$ is considered true unless p is true and q false. In Lewis' systems, the strict implication $p \Rightarrow q$ is considered true only if it is *impossible* that p is true and q false. Thus Lewis' strict implication can be defined from Russell and Whitehead's conjunction and negation signs and a new connective, ' $\Diamond$ ', of possibility.

$$p \Rightarrow q = \sim \Diamond (p \ \& \ \sim q)$$

Conversely, possibility can be defined from strict implication.

$$\Diamond p = \sim (p \Rightarrow \sim p)$$

Hence Lewis' project of finding the correct logical principles for his notion of strict implication is tantamount to that of finding the correct logical principles for possibility or, equivalently, those for necessity.

Lewis and Langford's *Symbolic Logic* (1932) describes five different axiom systems as candidates for the logic of strict implication. Much effort was expended in the first half of the century investigating these systems and variations of them. Even showing that all five are distinct (in the sense that they produce different classes of theorems) required considerable ingenuity. Fifteen years after the publication of *Symbolic Logic*, Carnap (1947) gave a non-axiomatic characterization of 'logical' necessity. '[Necessarily A]' is true, according to Carnap, if A is 'L-true', that is, if A is true in all state descriptions. (A 'state description' is a kind of canonical inventory of the primitive relations that hold and fail to hold of each sequence of individuals.) Thus Carnap can be seen as making precise the old idea that necessity is truth in all possible worlds. This idea is usually associated with Leibniz, but it is traced to Descartes by Curley (1984) and to Duns Scotus by Knuuttila (1982).

The logic determined by Carnap's interpretation turned out to be S5, the fifth of the Lewis and Langford systems. In the late 1950s and early 1960s, several authors proposed interpretations that refined and generalized Carnap's idea by introducing something like the accessibility relation described above (see [Kanger 1957](#); [Montague 1960](#); [Hintikka 1963](#); [Kripke 1963](#)). 'Kripke models', which are essentially the models described above, are the neatest formulation of this idea. Kripke and a continuing line of successors have shown that a great variety of modal

systems can be characterized by models of this kind. This enormously simplified the kinds of investigations of axiomatic systems mentioned above and opened new lines of research.

Kripke models were particularly fruitful in the study of modal logic in the broader sense. It had long been noted that the pairs ‘it will always/sometimes be the case that’, ‘it is obligatory/permitted that’ and ‘it is known/consistent with knowledge that’ exhibit logical behaviour similar to that of ‘necessarily/possibly’. The success of Kripke’s treatment of necessity encouraged analogous treatments of these other notions.

GA (‘It will always be the case that A ’) is true at time t if A is true at all times after t .

FA (‘It will sometimes be the case that A ’) is true at t if A is true at some times after t .

OA (‘It is obligatory that A ’) is true at world w if A is true at all worlds at which the obligations of w are discharged.

PA (‘It is permitted that A ’) is true at w if A is true at some such worlds.

KA (‘It is known that A ’) is true at world w if A is true in all worlds consistent with what is known at w .

The resulting systems are labelled ‘tense logic’, ‘deontic logic’ and ‘epistemic logic’ to distinguish them from the original ‘alethic’ modal systems for necessity. In the last two areas this account probably takes the analogy with necessity too far, but it still serves as a useful point of departure (see [Deontic logic](#); [Epistemic logic](#)).

Among other broadly modal systems that have received attention recently are the ‘dynamic logics’ or ‘logics of computation’ for reasoning about computer programs. Here worlds become computation states, which can be thought of as specifications of all the program variables at a particular time, and relative possibility becomes program accessibility, which holds between two states if a program can start in the first and terminate in the second. Such logics may be useful in verifying, without interminable testing, that a large program is ‘correct’, that is, that it does what it is supposed to do (see [Dynamic logics](#)).

The development of accessibility semantics led naturally to general questions about the classes of systems that can be characterized by various versions of it. Some of these are discussed below. Before the development of the accessibility interpretation, questions about the scope of axiomatic systems were often answered by devising suitable algebraic interpretations, and algebraic methods remain important tools for studying more general questions. Possible worlds semantics, however, seems less *ad hoc* than the algebraic methods. It is not clear whether they provide an analysis of necessity and possibility, or whether the notions that they incorporate, possible world and accessibility, are themselves to be analysed in terms of necessity and possibility. Either way, there is a close fit between the meanings of modal terms and their possible-worlds interpretations. The notion of possible world has proved useful in philosophical discussions on topics other than necessity and possibility: supervenience, causality, and the nature of propositions, properties and relations, to mention a few examples. The general utility of possible worlds has, in turn, inspired modal languages more expressive than the standard ‘box and diamond’ variety.

2 Propositional S5

As suggested above, S5 is the set of formulas provable from the axioms of classical propositional logic (PL) together with the axiom schemas K , T , 5 and $\Diamond A \leftrightarrow \neg \Box \neg A$ (henceforth $Df\Diamond$) and the rules *modus ponens* (MP) and necessitation (Nec). It is important to understand that Nec states that $\Box A$ is a theorem if A is, and not that $\Box A$ is true if A is. Given T and the replacement of equivalents, the latter condition would make $\Box A$ and A freely interchangeable, rendering modal logic pointless.

To identify a modal system in this way with its theorems and to refer to such systems as modal logics is to follow standard, though misleading, practices. One would expect a ‘logic’ to indicate which conclusions follow logically from which premises, and perhaps the deductions by which they so follow. Furthermore, although S5 and other modal systems are intended to represent formal truths, that is, sentences true in virtue of form, it is not clear that they are intended to represent *logical* truths, that is, sentences true in virtue of logical form, for it is not clear that necessity is a logical constant (see [Logical form](#); [Logical constants](#)). The first of these concerns can be somewhat reduced by stipulating that A follows from a set of formulas with respect to a logic L if, and only if, L contains a conditional whose antecedent is a conjunction of those formulas and whose consequent is A .

A model (for S5) is a pair (W, V) where W is a non-empty set (the possible worlds) and V is a function (the

valuation function) that assigns a truth-value (T or F) to each sentence letter and each possible world $w \in W$. If $M = (W, V)$ is a model and $w \in W$, the notion ‘ A is true at w in M ’ (written $(M, w) \models A$) is defined inductively, the key clauses being:

$(M, w) \models \Box B$ if, for all $v \in W$, $(M, v) \models B$; and
 $(M, w) \models \Diamond B$ if, for some $v \in W$ $(M, v) \models B$.

A formula is true in M if it is true at all possible worlds in M , and it is valid if it is true in all S5-models.

The ‘soundness theorem’ for S5 says that the semantics respects the logic in the sense that every formula in the logic is valid. The ‘completeness theorem’ states that every valid formula is in the logic. To say that S5 is both sound and complete with respect to the interpretation given is to say that validity and theoremhood coincide. Soundness is proved by an easy inductive argument appealing to the axiomatization: each axiom is observed to be valid and the two rules are shown to preserve validity. Completeness requires more ingenuity. A common approach is to adapt Henkin’s proof of the completeness of classical logic. (For details see Lemmon (1966) or Chellas (1980).) One can get an idea of the value of an interpretation and completeness theorem by trying, with and without them, to demonstrate that $\Box p \rightarrow \Box \Box p$ is a theorem of S5.

In S5, all ‘nesting’ or ‘iteration’ of modality can be eliminated. For example, if a formula has the form $M_1 \dots M_n A$ where each M_i is $\Box$, $\Diamond$ or negation, an equivalent formula can be obtained by deleting all the $\Box$ ’s and $\Diamond$ ’s except the innermost. Taking a modality to be a string of $\Box$ ’s, $\Diamond$ ’s and $\neg$ ’s, it follows that there are only six non-equivalent modalities in S5: the empty string, $\neg$, $\Box$, $\Diamond$, $\neg\Box$ and $\neg\Diamond$, corresponding to simple truth, falsity, necessity, possibility, non-necessity and impossibility.

One important schema of classical propositional logic that is not provable in S5 is extensionality:

(Ext) $(A \leftrightarrow A') \rightarrow (B \leftrightarrow B')$, where B' is the result of replacing an occurrence of the subformula A in B by A' .

For example, if p and q are both true at world w but only q is true at world v , then

$(p \leftrightarrow q) \rightarrow (\Box p \leftrightarrow \Box q)$

is false at w . By completeness it follows that

$(p \leftrightarrow q) \rightarrow (\Box p \leftrightarrow \Box q)$

is not a theorem of S5. Ext says that replacement of one subformula by another of the same truth-value will not affect the truth-value of the whole. Its failure is often viewed as characteristic of modal systems in general. Note, however, that provably equivalent formulas can be substituted for each other.

S5 is a relatively strong modal system. Its only extensions are the trivial logic, containing all instances of $\Box A \leftrightarrow A$ and, for every natural number n , the ‘ n -possibility logic’, containing all instances of

$(\Diamond A_1 \wedge \dots \wedge \Diamond A_{n+1}) \rightarrow \text{Dis}$,

where Dis is the disjunction of all formulas $\Diamond(A_i \wedge A_j)$ such that $i < j \leq n + 1$. (n -possibility logic is complete with respect to the class of all models with at most n possible worlds.)

The formulas of (propositional) S5 correspond to formulas of classical monadic predicate logic in one variable. For example, $\Diamond(p_1 \wedge \Box p_2)$ corresponds to $\exists x(P_1 x \wedge \forall x P_2 x)$. Since decision procedures for monadic predicate logic are known, this correspondence allows one to determine effectively whether a formula is in S5 by testing the corresponding monadic formula for validity. A more direct proof of decidability rests on the result that S5 has the ‘finite model’ property: every non-theorem is false in some model with finitely many worlds. To test whether A is in S5, one checks first whether A is provable in one step or falsifiable in a one-world model, then whether it is provable in two steps or falsifiable in a two-world model, and so on. Each step can be completed in finite time. By the finite model property, there is some step n at which the process yields the desired answer.

3 Philosophical questions about S5

The most important philosophical question about S5 is whether it captures the inferences and truths it was intended

to. This may depend, of course, on the kind of necessity that ‘ $\Box$ ’ is supposed to represent. On the usual view this is broadly logical or metaphysical necessity. Truths necessary in this sense include those true in virtue of logical form (the logical truths), those true in virtue of meaning (the analytic truths) and a more problematic category of those true in virtue of the basic nature of things. The last category has been said to contain truths of mathematics, the proposition that water is H_2O , and the proposition that Queen Elizabeth came from the egg and sperm that she did (see **Essentialism**). All these examples are controversial, but those who argue whether particular propositions are metaphysically necessary may nevertheless share a common conception of what it is to be metaphysically necessary. Furthermore, examples of propositions that lack metaphysical necessity seem uncontroversial: ‘Napoleon invaded Russia’, ‘Asbestos is carcinogenic’, ‘Paris is the capital of France’.

The question of whether S5 (or any other system) is the right ‘logic’ for metaphysical necessity can be divided into two parts, corresponding to the two parts of the completeness theorem. Say that S5 is ‘correct’ if every theorem represents a formal truth about metaphysical necessity. Say that it is ‘adequate’ if every formal truth (with ‘and’, ‘not’, ‘or’ and ‘if’ as well as ‘necessarily’ and ‘possibly’) is represented by a theorem. Correctness and adequacy, then, are philosophical counterparts of soundness and completeness.

Correctness can be established by an argument similar to that for soundness: first show that the axioms represent formal truths and then that the rules transform formulas representing formal truths into formulas representing formal truths. Instances of axiom schema *T*, for example, clearly represent formal truths. (‘If necessarily $87 + 25 = 112$ then $87 + 25 = 112$ ’ is true in virtue of its form.) The difficult cases are axiom schema 5 and the rule Nec. For the former case we need to establish that sentences of the form ‘If possibly *S* then necessarily possibly *S*’ are true in virtue of their form. One argument is as follows. To say possibly *S* is to say that ‘Not *S*’ does not follow logically or analytically from a description of the natures of things or from logical truths or analytic truths, that is, that *S* is logically and analytically consistent with the basic natures of things. But a proposition that something is consistent with logical laws, meanings and basic natures in this way is true or false in virtue of those logical laws, meanings and basic natures. So ‘Possibly *S*’, if true, is necessarily true. And since this argument does not appeal to *S*, the conditional is true in virtue of its form. An argument for the rule Nec might go as follows. Once it is established that all the axioms represent formal truths and that MP preserves formal truth, we know that everything proved without Nec represents a formal truth. Since formal truths are either logical or analytic, they are necessarily true. Since the argument that any of these sentences *S* is necessary relies only on the form of *S*, necessarily *S* is true in virtue of its form. This establishes that the first application of Nec results in formulas that represent formal truths. But if a sentence *S* proved with one application of Nec is shown to be a formal, and hence necessary, truth by an argument that appeals only to *S*’s form, then ‘Necessarily *S*’ must be true in virtue of its form. In this way, the argument can be extended to any subsequent application of Nec.

Adequacy might be established indirectly. Suppose that S5 were not adequate for metaphysical necessity. Then there would be some formal truth *S*, with ‘necessarily’, ‘possibly’, ‘and’ and so on, represented by a formula *A* that is not a theorem of S5. We should, then, be able to ‘improve’ the adequacy of S5 by adding *A* as an axiom. Furthermore, all the substitution instances of *A* will represent sentences with the form of *S*, which, if *S* is a formal truth, will also be formal truths. So we should be able to add all the substitution instances of *A* as well, obtaining an extension of S5. But, as was noted above, the extensions of S5 must contain either $\Box A \leftrightarrow A$ or, for some *n*, $(\Diamond A_1 \wedge \dots \wedge \Diamond A_{n+1}) \rightarrow \text{Dis}$. All these are incorrect for metaphysical necessity.

4 Quantified S5

Consider a language obtained by adding ‘ $\Box$ ’ and ‘ $\Diamond$ ’ and a special predicate *E* of existence to a version of predicate logic with predicates $P_1, P_2, \dots$, individual constants, $t_1, t_2, \dots$, and the identity sign, ‘=’. A model is a triple (W, D, V) where *W* is a non-empty set (the possible worlds), *D* is a function that assigns to each $w \in W$ a set D_w (the domain of *w*) and *V* is a function that assigns to each constant *t* a member of the union $\cup D$ of the sets D_w (the possible object ‘denoted’ by t_i) and to each world and *n*-place predicate an *n*-ary relation on $\cup D$. A definition of truth at a world is obtained from the previous one by replacing the base clause and adding clauses for ‘ $\forall$ ’, ‘ $\exists$ ’, ‘=’ and *E*. The quantifier clauses state that, for example, $\forall x Px$ is true at *w* if, for every *d* in D_w , *Pt* is true at *w* when *t* denotes *d*, and that $\exists x Px$ is true at *w* if, for some *d* in D_w , *Pt* is true at *w* when *t* denotes *d*. $s = t$ is true at *w* if *s* and *t* denote the same possible objects. *Es* is true at *w* if $V(s) \in D_w$.

This interpretation reflects several choices. First, constants are ‘objectual’, that is, V assigns a possible object directly to each constant. $\Box Pcd$ is true at w if the possible objects $V(c)$ and $V(d)$ are related, at every world w , by $V(P, w)$. Thus $\Box Pcd$ expresses a *de re* necessity: it asserts that particular objects, independently of their descriptions, are necessarily related. This treatment of constants makes the schemas

$$(c = d) \rightarrow \Box(c = d) \quad (\text{‘necessary identity’}) \text{ and} \\ \neg(c = d) \rightarrow \Box\neg(c = d) \quad (\text{‘necessary difference’})$$

valid. Kripke has influentially argued that this is appropriate if c and d represent proper names of natural language, but generally inappropriate if they represent definite descriptions (see **Proper names**). The alternative to treating constants objectually is to allow their denotations to vary from world to world. An (individual) concept is a function from worlds to individuals. For example, the concept ‘first person to reach the South Pole’ might assign Amundsen to this world, Scott to the possible world in which Scott wins his race with Amundsen, and nothing to possible worlds in which the Pole is never (or always) occupied by people. We can regard a constant with a non-objectual interpretation as denoting a concept rather than an individual. On this ‘conceptual’ interpretation, necessary identity and necessary difference are not valid. Quantification may be treated conceptually as well: $\forall xA$ is declared true at w if every individual concept assigns to w an object to which A applies at w . In that case, quantified necessary identity and difference formulas,

$$\forall x\forall y(x = y \rightarrow \Box x = y) \text{ and} \\ \forall x\forall y(\neg(x = y) \rightarrow \Box\neg(x = y)),$$

lose their validity as well. The quantification described above, by contrast, is objectual.

Second, quantification is ‘actualist’. The truth of $\forall xPx$ at w requires only that the objects in D_w have the appropriate property. It follows that the Barcan formula

$$\forall x\Box Px \rightarrow \Box\forall xPx$$

and its converse

$$\Box\forall xPx \rightarrow \forall x\Box Px$$

are both invalid. Suppose, for example, that for some world w , P is a predicate that holds everywhere of just the objects that exist in w , and that D_u contains something not in D_w . Then at w , ‘Everything is necessarily P ’ is true while ‘Necessarily, everything is P ’ is false. Conversely, suppose that, at every world w , P holds of just the objects that exist at w , and that D_v contains something not in D_u . Then at v ‘Necessarily everything is P ’ is true, while ‘Everything is necessarily P ’ is false. More generally, $\forall x\Box\forall y\Box A$, $\Box\forall x\forall y\Box A$, $\Box\forall x\Box\forall yA$, and $\Box\forall x\Box\forall y\Box A$ are logically distinct.

If quantification is ‘possibilist’, that is, if $\forall xA$ means that A is true of all objects in $\cup D$, then these formulas are all equivalent and the Barcan formula and its converse are valid. The distinction between actualist and possibilist quantification is significant only because the models defined above have domains that vary from world to world. If one stipulates ‘constant domains’, that is, that $D_u = D_v$ for all worlds u and v , then the possible objects are just the actual ones, and the distinction collapses. Barcan and its converse are again valid.

Third, predicates can be truthfully applied to constants denoting non-actual objects. Since quantification is actualist, this implies that the classical theorems $Pc \rightarrow \exists xPx$ and $\forall xPx \rightarrow Pc$ are not valid. These principles can be retained by insisting that the application of predicates to constants denoting non-actuals is always false (defining models so that $V(P_i, w)$ is a relation on D_w). But to do so would be to adopt a kind of atomism according to which the properties and relations expressed by atomic formulas with free variables had a special status. Furthermore, although it would save the particular classical theorems above, some of their substitution instances, such as $\neg Ex \rightarrow \exists x\neg Ex$, would still fail. Another alternative is to stipulate that formulas in which predicates are applied to constants denoting non-existents lack truth-value. If validity is taken as ‘false in no model’, both the formulas and their instances are saved. On the other hand, that approach would seem to make $\Box Ec$ valid (if necessity is ‘nowhere false’) or to make $\Box(Fc \vee \neg Fc)$ invalid (if necessity is ‘everywhere true’).

Finally, domains are permitted to be empty. This implies that

$\Diamond \exists x(Fx \vee \neg Fx)$, $\Box \exists x(Fx \vee \neg Fx)$ and $\exists x(Fx \vee \neg Fx)$

are all invalid. If one regards the first as expressing a formal truth, one can require that some world have non-empty domain; if the second, that all worlds do. If one regards the third as expressing a formal truth one can require that each model specify - in addition to the possible worlds, domains, and valuation - a particular possible world (the actual) that has a non-empty domain. Truth in a model is then redefined as truth at the actual world of the model.

5 Weaker systems

A Kripke model is a triple (W, R, V) , where W and V are as in §2, and R (accessibility) is a binary relation on W . Truth at a world is defined as before except for the ‘ $\Box$ ’ and ‘ $\Diamond$ ’ clauses:

$(M, w) \models \Box A$ if, for all $v \in W$ such that wRv ,
 $(M, v) \models A$
 $(M, w) \models \Diamond A$ if, for some $v \in W$ such that wRv ,
 $(M, v) \models A$

Truth in a model and validity are defined as before. The formulas valid in this sense comprise the logic K. K can be axiomatized by the schemas PL, K and Df $\Diamond$ and the rules MP and Nec. Since it lacks the schema $\Box A \rightarrow A$, K is not adequate for necessity under any construal. It occupies an important position in modal logic in the broader sense, however, because many well-known modal systems are simple extensions of it. The systems in the leftmost column of the table, for example, are obtained by adding the schemas in the middle column to PL, K and Df $\Diamond$ and the rules MP and Nec.

System	Characteristic axioms	Conditions on R
D	$D: \Box A \rightarrow \Diamond A$	seriality: $\forall x \exists y Rxy$
T	T	reflexivity: $\forall x Rxx$
S4	$T\ 4: \Box A \rightarrow \Box \Box A$	reflexivity transitivity: $\forall x \forall y \forall z (Rxy \ \& \ Ryz \rightarrow Rxz)$
S4.3	$T, 4$ $H: \Diamond A \ \& \ \Diamond B \rightarrow \Diamond(A \wedge \Diamond B) \vee \Diamond(\Diamond A \wedge B) \vee \Diamond(A \wedge B)$	reflexivity, transitivity connectedness: $\forall x \forall y (Rxy \vee Ryx)$
GL	$W: \Box(\Box A \rightarrow A) \rightarrow \Box A$	transitivity no infinite chains: $Rx_1x_2 \ \& \ Rx_2x_3 \ \& \ \dots \rightarrow \exists i(x_i = x_{i+1})$

The schema D (see the table) is formally true when ‘ $\Box$ ’ and ‘ $\Diamond$ ’ are read ‘it is obligatory that’ and ‘it is permitted that’, and the system D is known as the standard deontic logic. T was one of the earliest modal logics to be characterized precisely (see Feys 1937, 1938) and it seems to be the weakest logic in which ‘ $\Box$ ’ can plausibly be regarded as representing a reading of ‘it is necessary that’. If, as some have suggested, the remaining S5 axioms are not correct for physical necessity, T would be a plausible candidate for the logic of that notion. S4 was the fourth of the Lewis systems. Gödel (1933) gave it a characterization like the one above and showed it to be intertranslatable with intuitionistic propositional logic. S4.3 (so named, in part, because it is intermediate in strength between S4 and S5) is correct and adequate for a reading of ‘ $\Box$ ’ as ‘it is and always will be the case that’ if time is assumed to be linear (see Tense and temporal logic). If sentence letters represent statements about numbers and ‘ $\Box$ ’ is interpreted as ‘provable in arithmetic’ then the system GL contains exactly the formulas that are themselves provable in arithmetic (where a statement about provability is ‘provable’ if its Gödel number is the number of a theorem of arithmetic - see Gödel’s theorems). By adding all instances of $\Box A \rightarrow A$ to the theorems of GL (without allowing any new applications of the rule Nec) one obtains the system GLS, which may be regarded as the logic of arithmetic provability. Whereas GL comprises the arithmetically provable formulas about provability, GLS comprises the truths about provability (see Provability logic).

Each of these logics can also be characterized semantically, just as we have done with K. A ‘D-model’ is a Kripke model whose accessibility relation is ‘serial’, that is, every world is related to some world or other, and models appropriate for each of the other systems can be similarly defined by the conditions in the table. In each case, soundness and completeness results like those for S5 sketched above can be given.

Not all broadly ‘modal’ systems can plausibly be interpreted by Kripke models. Suppose $\Box A$ is read ‘usually A’. There is no relation on times such that ‘usually A’ is true now if A is true at all related times. Rather, the truth of ‘usually A’ depends on the number, and perhaps the distribution, of all the times at which A is true. This suggests a more general kind of modal semantics, one formulation of which is the ‘neighbourhood semantics’. A ‘neighbourhood model’ is a triple (W, R, V) , where W and V are as before, and R is a relation between worlds and sets of worlds (the ‘neighbourhoods’ of those worlds, although there is no requirement that the neighbourhood of a world contain the world itself, or even that it be non-empty). The definition of truth at a world is as before except that $\Box A$ is true at w just in case w is related to the ‘truth-set’ of A , that is, the set of worlds at which A is true; and $\Diamond A$ is true at w if w is unrelated to the ‘falsity-set’ of A .

The modal system determined by the set of all such models is the system E, axiomatized by the schemas PL and Df $\Diamond$ and rules MP and ‘equivalents’ (RE): ‘If $A \leftrightarrow B$ is provable, so is $\Box A \leftrightarrow \Box B$ ’. Like K, E provides a convenient base from which a variety of systems of interest can be constructed, rather than a characterization of the formal truths for some particular reading of ‘ $\Box$ ’. This idea can be carried even further. Operator logic (OL) is the system in the language with ‘ $\Box$ ’ that is axiomatized just by PL (see Kuhn 1981). The modal operators of S5, K, E and OL can be regarded as being successively more schematic and having successively less content, the theorems of OL being the modal sentences that are true in virtue of their logical form, and the stronger modal systems being theories of operators based on that logic.

6 General results

Much of the contemporary study of modal logic is directed, not towards investigating any particular one of the systems described above in more detail, but towards a general understanding of classes of such systems. The modal logics, as defined above, form a lattice structure. A Kripke model $M = (W, R, V)$ can be viewed as the addition of a valuation V to the ‘Kripke frame’ (W, R) . M is then said to be ‘based on’ (W, R) . Similarly, the neighbourhood model (W, N, V) is based on the neighbourhood frame (W, N) . A formula is valid in a frame if it is true in all models based on the frame, and it is valid in a class of frames if it is valid in all frames of the class. A ‘frame for L ’ is a frame in which all the theorems of L are valid. The set of formulas valid in a class of frames is a logic, the logic determined by that class. A logic is sound for a class of frames if every formula in the logic is valid in the class; it is sufficient for the class if every formula valid in the class is a member of the logic; it is complete (for the class) if it is sound and sufficient.

The logic determined by a class of frames is, by definition, complete for that class. It might also be complete for other classes. K, for example, is determined by (and therefore complete for) the irreflexive Kripke frames as well as all Kripke frames. Fine (1974) and Thomason (1974) independently showed that there are finitely axiomatizable extensions of K that are not determined by any class of Kripke frames. Such logics are incomplete: formulas true in every frame that verifies the axioms are unprovable. It is now known that the incompleteness phenomenon is widespread. For every extension L of the logic T there are uncountably many incomplete logics whose frames are exactly the frames for L . (This result holds for both Kripke frames and neighbourhood frames; see Benton 1985.) The incomplete logics that have been exhibited in the literature are generally complex and *ad hoc*, but one simple example is obtained from GL by replacing the first conditional of the schema W by a biconditional. In the other direction, results of Bull (1966) and Fine (1971) imply that every extension of S4.3 that admits the necessitation rule is complete and decidable.

Much work has also been done on the correspondences between modal and classical formulas. The schema T corresponds to the classical formula $\forall x Rxx$ in the sense that the frames for the former are just the first-order models for the latter. Most modal schemas that have arisen naturally in philosophical discussion correspond similarly to first-order formulas. The schemas in the second column of the table, for example, correspond to formulas in the third column. The McKinsey schema, $\Box \Diamond A \rightarrow \Diamond \Box A$, on the other hand, corresponds to no first-order formula (see van Benthem 1975; Goldblatt 1975), and no modal schema corresponds to irreflexivity:

$\forall x \neg Rxx$. (The latter fact follows from the remark above that K itself is complete for the irreflexive frames.)

The general study of modal logic encompasses a variety of other topics, including: model theory (transformations of frames and models may preserve truth-values of classes of formulas); boundary investigations (some logics have properties that all their extensions lack); expressive power (some classes of modal connectives can be defined from a few representatives); and connections with non-modal logics (such as that between S4 and intuitionistic logic). The completeness and correspondence investigations discussed above, however, have formed the core of contemporary investigations.

See also: [Logical and mathematical terms, glossary of](#); [Modal logic, philosophical issues in](#)

STEVEN T. KUHN

References and further reading

- Aristotle** (c.mid 4th century BC) *De Interpretatione, Prior Analytics and Metaphysics*, trans. and ed. J. Barnes, in *The Revised Oxford Translation*, vols 1 and 2, Princeton, NJ: Princeton University Press, 1984.(Referred to in §1.)
- Bentham, J.F. van** (1975) 'Some Connections Between Modal and Elementary Logic', in S. Kanger (ed.) *Proceedings of the Third Scandinavian Logic Symposium*, Amsterdam: North Holland.(Referred to in §6.)
- Benton, R.** (1985) 'General Modal Incompleteness and Finite Axiomatizability', Ph.D. dissertation, University of Michigan. (Referred to in §6.)
- Boolos, G.** (1993) *The Logic of Provability*, Cambridge: Cambridge University Press.(Thorough and readable exploration of the arithmetic provability interpretation of $\Box$ discussed in §5.)
- Bull, R.A.** (1966) 'That all Normal Extensions of S4.3 Have the Finite Model Property', *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik* 12: 341-44.(Referred to in §6.)
- Carnap, R.** (1947) *Meaning and Necessity: A Study in Semantics and Modal Logic*, Chicago, IL: University of Chicago Press; 2nd, enlarged edn, 1956.(Referred to in §1.)
- Chellas, B.F.** (1980) *Modal Logic: An Introduction*, Cambridge: Cambridge University Press. (Referred to in §2. Careful and clear introductory text on propositional modal logic, emphasizing completeness and decidability.)
- Curley, E.M.** (1984) 'Descartes on the Creation of the Eternal Truths', *Philosophical Review* 93: 569-97.(Referred to in §1.)
- Feys, R.** (1937, 1938) 'Les logiques nouvelles des modalités', *Revue Néoscholastique de Philosophie* 40: 517-53, 41: 217-52.(Referred to in §5.)
- Fine, K.** (1971) 'The Logics Containing S4.3', *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik* 17: 371-6.(Referred to in §6.)
- Fine, K.** (1974) 'An Incomplete Logic Containing S4', *Theoria* 40: 23-9.(Referred to in §6.)
- Fine, K. and Kuhn, S.** (forthcoming) *Modal Logic*.(Text and survey that is attentive to philosophical concerns, including the issues of correctness and adequacy discussed in §3.)
- Gabbay, D. and Guenther, F.** (eds) (1984) *Handbook of Philosophical Logic*, vol. 2, *Extensions of Classical Logic*, Dordrecht: Kluwer.(The articles by Bull and Segerberg, van Bentham, Garson and Cocchiarella constitute an advanced survey of central topics. The volume also contains excellent articles on modality and time by R. Thomason, modality and self-reference by Smorynski, and others on topics somewhat more peripheral to the present one.)
- Gödel, K.** (1933) 'Eine Interpretation des intuitionistischen Aussagenkalküls', *Ergebnisse eines Mathematischen Kolloquiums* 4: 39-40; trans. 'An Interpretation of the Intuitionistic Propositional Calculus', in *Collected Works*, ed. S. Feferman, J.W. Dawson, W. Goldfarb, C.D. Parsons and R.M. Solovay, Oxford and New York: Oxford University Press, 1995, vol. 3, 296-302.(Referred to in §5.)
- Goldblatt, R.** (1975) 'First-order Definability in Modal Logic', *Journal of Symbolic Logic* 40: 35-40. (Referred to in §6.)
- Hintikka, J.** (1963) 'The Modes of Modality', *Acta Philosophica Fennica* 16: 65-81; repr. in *Models for Modalities*, Dordrecht: Kluwer, 1969.(Referred to in §1.)
- Hughes, G.E. and Cresswell, M.J.** (1968) *An Introduction to Modal Logic*, London: Routledge; revised edn, *A New Introduction to Modal Logic*, London: Routledge, 1996.(The classic modern textbook; still useful.)
- Hughes, G.E. and Cresswell, M.J.** (1984) *Companion to Modal Logic*, London: Routledge.(A sequel to the previous book that incorporates subsequent developments in the field, including an incompleteness result.)
- Kanger, S.** (1957) *Provability in Logic*, Stockholm: Almqvist & Wicksell.(Referred to in §1.)
- Knuuttila, S.** (1982) 'Modal Logic', in N. Kretzmann, A. Kenny and J. Pinborg (eds) *Cambridge History of Later*

- Medieval Philosophy*, Cambridge: Cambridge University Press.(Referred to in §1. Brief survey of some medieval thought on modality, including that of Boethius, Scotus and Buridan.)
- Knuuttila, S.** (1993) *Modalities in Medieval Philosophy*, London and New York: Routledge.(More comprehensive survey than Knuuttila (1982), with an extensive bibliography.)
- Kripke, S.A.** (1963) 'Semantical Analysis of Modal Logic I: Normal Modal Propositional Calculi', *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik* 9: 67-96.(Referred to in §1.)
- Kuhn, S.** (1981) 'Constants, Logical Symbols and Operator Logic', *Journal of Philosophy* 78: 487-99. (Referred to in §5.)
- Lemmon, E.J.** (1966) *An Introduction to Modal Logic*, ed. K. Segerberg, Oxford: Blackwell, 1977. (Referred to in §1. Draft of initial chapters of a planned collaboration with D. Scott. In fewer than 100 pages, it includes a historical survey and many applications of the Henkin completeness method.)
- Lewis, C.I. and Langford, C.H.** (1932) *Symbolic Logic*, New York: Century.(Appendix II, which was written by Lewis, lays out the axiom systems S1-S5 for strict implication that are mentioned in §1.)
- Mints, G.** (1992) *A Short Introduction to Modal Logic*, Stanford, CA: Center for the Study of Language and Information, Stanford University.(A concise treatment of the systems S5, S4 and T that emphasizes refutation procedures inspired by Gentzen's sequent calculi, rather than the axiomatic characterizations presented here.)
- Montague, R.** (1960) 'Logical Necessity, Physical Necessity, Ethics and Quantifiers', *Inquiry* 4: 259- 69; repr. in *Formal Philosophy*, New Haven, CT: Yale University Press, 1974.(Referred to in §1.)
- Thomason, S.K.** (1974) 'An Incompleteness Theorem in Modal Logic', *Theoria* 40: 30-4.(Referred to in §6.)

Modal logic, philosophical issues in

In reasoning we often use words such as 'necessarily', 'possibly', 'can', 'could', 'must' and so on. For example, if we know that an argument is valid, then we know that it is necessarily true that if the premises are true, then the conclusion is true. Modal logic starts with such modal words and the inferences involving them. The exploration of these inferences has led to a variety of formal systems, and their interpretation is now most often built on the concept of a possible world.

Standard non-modal logic shows us how to understand logical words such as 'not', 'and' and 'or', which are truth-functional. The modal concepts are not truth-functional: knowing that p is true (and what 'necessarily' means) does not automatically enable one to determine whether 'Necessarily p ' is true. ('It is necessary that all people have been people' is true, but 'It is necessary that no English monarch was born in Montana' is false, even though the simpler constituents - 'All people have been people' and 'No English monarch was born in Montana' - are both true.)

The study of modal logic has helped in the understanding of many other contexts for sentences that are not truth-functional, such as 'ought' ('It ought to be the case that p ') and 'believes' ('Alice believes that p '); and also in the consideration of the interaction between quantifiers and non-truth-functional contexts. In fact, much work in modern semantics has benefited from the extension of modal semantics introduced by Richard Montague in beginning the development of a systematic semantics for natural language.

The framework of possible worlds developed for modal logic has been fruitful in the analysis of many concepts. For example, by introducing the concept of relative possibility, Kripke showed how to model a variety of modal systems: a proposition is necessarily true at a possible world w if and only if it is true at every world that is possible relative to w . To achieve a better analysis of statements of ability, Mark Brown adapted the framework by modelling actions with sets of possible outcomes. John has the ability to hit the bull's-eye reliably if there is some action of John's such that every possible outcome of that action includes John's hitting the bull's-eye.

Modal logic and its semantics also raise many puzzles. What makes a modal claim true? How do we tell what is possible and what is necessary? Are there any possible things that do not exist (and what could that mean anyway)? Does the use of modal logic involve a commitment to essentialism? How can an individual exist in many different possible worlds?

1 What makes a modal claim true?

Alice ate vegetable soup today. What makes that true? Alice's activities and their relationship to vegetable soup. Alice *could have* eaten cream of tomato soup. What makes that true? This is less clear. No activity of Alice's seems to make this statement about possibilities true.

If we ask what makes something possible, the context of the question sometimes sets a standard: it was possible for Alice to have eaten cream of tomato soup because it was on the menu, there was plenty of it available, there was nothing barring her from ordering it and nothing standing in the way of the waiter's bringing it to her. Yet it seems that there is a more absolute sense of possibility, according to which it was at least logically possible for Alice to have eaten cream of rutabaga soup (even if no one had ever made such a thing), because no logical or conceptual barrier prevents its existence or her eating it.

In contrast, Alice could not eat cream of tomato soup without eating it, and she could not have vegetable soup that involved no vegetables at any stage of its preparation. For logical or conceptual reasons, these are not activities open to Alice. Not every description picks out something that is possible.

Once we get a view of this broader notion of possibility, it seems that possibility can get a negative characterization. A description of an event or state of affairs describes something possible if no logical or conceptual consideration rules out what it describes. It is possible if it is not impossible; and the impossible is whatever is ruled out by logical or conceptual considerations. If a proposition is necessary, then its negation is impossible. The proposition

If Alice eats tomato soup, then Alice eats tomato soup

is necessary; and so a negation of that, a proposition equivalent to
Alice eats tomato soup without eating tomato soup,

is impossible. It is necessary that

All vegetable soup involves vegetables in its preparation;

so it is impossible that

Something is vegetable soup and does not involve vegetables in its preparation.

Understanding what is impossible will follow upon an understanding of what is necessary.

All of this suggests that we can understand what makes something possible if we can get a general understanding of what makes something a necessary truth. An event or state of affairs p is not possible if and only if p is impossible; p is impossible if and only if the negation of p , $\sim p$, is necessary. Ordinarily the box symbol is used before a sentence to stand for necessity (' $\Box p$ ' stands for 'It is necessary that p ') and the diamond is used to stand for possibility (' $\Diamond p$ ' stands for 'It is possible that p '), and we have just taken note of these related equivalences:

$\Diamond p$ iff $\sim \Box \sim p$.

$\sim \Diamond p$ iff $\Box \sim p$.

Because of these logical relationships between possibility and necessity, we can understand what makes something possible by developing an understanding of what makes something necessary. A state of affairs is possible if and only if its negation is not necessary. So what makes something a necessary truth? Our examples so far have consisted of logical and conceptual truths, but it seems that the category of necessary truths is broader. My pet cat could not exist without being an animal; it is necessary that if she exists, then she is an animal. I could not have existed if a certain sperm and ovum had never existed; it is necessary that if I exist, then those cells existed. No right hand can precisely fill the space taken up by any left hand; it is necessary that if a space S is filled by a left hand, then S cannot be precisely filled by any right hand. Logic and definitions do not seem to suffice as the source of the necessity of these truths.

In exploring logical relationships, logicians have introduced the concept of a possible world. This has been of enormous benefit in understanding logical relationships among modal claims. A statement is a necessary truth if and only if it is true at every possible world; a statement is possibly true if and only if it is true at at least one possible world. Thus our symbols can be connected with a systematic semantics for modal language in which possible worlds are reference points:

Any sentence of the form ' $\Box p$ ' is true iff p is true at every possible world.

Any sentence of the form ' $\Diamond p$ ' is true iff p is true at at least one possible world.

Kripke's elaboration of this, employing the notion of relative possibility, has been very important in exploring a range of modal systems that give a range of different results concerning the logical relationships among modal sentences (see [Kripke, S.A.; Modal logic](#)). With further elaboration, introducing a 'closeness' relation, this framework has been extended (see [Lewis 1973](#), [Stalnaker 1968](#)) to provide a better semantics for counterfactuals: roughly, a counterfactual 'If it were the case that p , then it would be the case that q ' is true at w if and only if q is true at all of the closest worlds to w in which p is true.

Such systematizations seem to give us an account of the necessity and possibility of propositions that can help us to see more clearly what is and what is not possible. This account, however, is not an ultimate answer to the question of what makes something possible, because there is no independent account of what a possible world is. We are left with the question of what makes it the case that Alice's eating rutabaga soup is included in some possible worlds, whereas my cat's being a plant is not. At a possible world, every necessary truth must be true, every state of affairs which obtains must be possible, and the possible world must meet some maximality condition (for example, that every proposition or its negation must be true at that world) - but then judging what is included in a possible world takes us back to the question of judging what is possible. So we have a circle of concepts. The

exploration of such a circle of concepts can be very enlightening concerning their logical relationships, and it can help us to make more consistent judgments about the application of the concepts, but it cannot give an independent answer to the question of what makes a modal statement true (see [Possible worlds](#); [Semantics, possible worlds](#)).

Thus we are led to a root philosophical problem about modality, the problem of giving a general account of what makes modal claims and related claims true - or, which amounts to the same problem, the problem of saying in general what makes a set of propositions true of a possible world (or what makes a set of states of affairs constitute a possible world). Any answer that is non-modal seems inadequate, and any answer that is modal merely enlarges the circle. Perhaps the kind of understanding that we get by enlarging the circle and exploring the relationships among modal concepts is the best we can hope for (see [Blackburn 1993](#)).

2 The merely possible

Even when we work within the circle of modal concepts, attempting to get a coherent, systematic semantics for modal claims, we come upon many puzzles. Understanding the relationship between the actual and the possible is at the root of much of the difficulty.

It seems clear that there could have been more carrier pigeons. If people had not shot so many of them, they would be flying now, perhaps in flocks so large that they would darken the skies. In saying this - in using the word 'they' to refer to some carrier pigeons that *would be* flying now - I seem to be referring to carrier pigeons that do not exist. Are there carrier pigeons that do not exist, that is, some nonexistent, merely possible carrier pigeons? We should hope not. Yet in a systematic treatment of modality, it is difficult to avoid the problem of requiring the existence of merely possible objects, and various strategies have been employed to deal with this problem. (Not everyone regards a commitment to nonexistent objects as a problem; see [Meinong, A.](#))

One strategy is to note a distinction between existence and actuality. Some things - possibilia - exist without being actual, and the merely possible carrier pigeons are among them. In fact, it might be said that if possible worlds exist without being actual, why should we regard it as a new problem to accept the existence of the things that make up those worlds, such as possible carrier pigeons? If the possible worlds exist, it seems that the things in them should exist as well. The use of modal language commits us to both possible worlds and possibilia, according to this view.

The view that there are some merely possible objects, just as there are merely possible worlds, seems grounded in a puzzle, however, and we can separate that view from the thesis that other possible worlds exist. A distinction between existence and actuality makes sense for possible worlds, on most conceptions, because a possible world is really an abstract representation; a way things might be, rather than a way in which they are. (The conception for which David [Lewis](#) argues is an important exception to this; see [Lewis 1986](#).) Thus we can say that there are worlds very much like this one except that Quine becomes a full-time geographer and never becomes a philosopher. There are such ways for things to be, but they are not actualized: things are not that way. This distinction between existing and being actualized holds among worlds as it does for stories or dreams, because a world is a way things might be; a maximal state of affairs or a maximally consistent set of propositions. This distinction does not carry over to individuals such as Quine and carrier pigeons. If non-actual carrier pigeons are carrier pigeons, then they are not the sort of thing that can exist without being actualized. This is the puzzle of possibilia, because we cannot see how any carrier pigeon can exist without being actual (in the way that a story can exist without being actual), and the idea that possible worlds and possible objects are on an existential par seems unacceptable. Possibilia cannot simply ride in on the coat tails of possible worlds.

One way out of this is to interpret the semantics as being about something more abstract than individuals - concepts or individual essences (see [Plantinga 1974](#), for example). Instead of using the ordinary concept of satisfaction in the semantics, we can use a concept 'satisfaction^{*}', saying that an individual essence *e* satisfies* 'x is a geographer' at a world *w* if and only if *e* is an essence of something that would be a geographer if that world, *w*, were actual. These individual essences can then be things that exist even when they are not actualized and that are actualized by the same individual at every world at which they are actualized. The individual essences of many carrier pigeons exist without being actualized. That is why it is true that other carrier pigeons could have existed.

On this view, individual essences rather than individuals are components of propositions. This allows the

propositions to exist even when the individuals do not exist. This conflicts with David Kaplan's view that a typical singular proposition is contingently existent, existing only in those possible worlds in which the individual that it is about exists (see [Kaplan 1989](#); [Demonstratives and indexicals §2](#)). But it need not conflict with his more central views: that singular propositions do not contain concepts that can be used to pick out the individual the proposition is about; and that the propositional component is determined by the individual rather than the other way around. This would only mean that possible worlds in which an individual does not exist provide no resources for singling out that individual's essence. Nothing in this approach conflicts with that.

The approach in terms of essences leaves puzzles, too. What are these things? How many of them are there? These questions are not easy to answer, and so we are left with some discomfort in understanding modal claims no matter how we try to deal with the problem of possibilities.

3 Quantification, modality and essentialism

Many of those who have considered the concept of necessity have regarded it as a logical concept, applying to sentences in virtue of their form. An idea like one of the following would be typical.

A sentence of the form ' $\Box p$ ' is true iff p is a logical truth.

A sentence of the form ' $\Box p$ ' is true iff p is analytic.

W.V. [Quine](#) has expressed grave misgivings about this at two levels. First, he questions the intelligibility of the notion of analyticity, and at times seems even to question the concept of logical truth. Thus he will reject any notion of necessity construed in this way. Quine's misgivings about logical truth and analyticity are explored elsewhere (see [Analyticity](#)).

Quine ([1961](#), [1966](#)) has a further objection to the intermingling of quantifiers and modalities. Even if one grants a concept of analyticity or logical truth to serve as the basis for the concept of necessity, there are two problems for quantified modal logic: the problem of quantifying into modal contexts; and the related problem of intersubstituting coreferential names.

We can see the problem of quantifying in more easily if we consider first the problem of quantifying into truth-functional contexts. This is a real problem, because we ordinarily give an account of truth-functional connectives in terms of the truth-values of the constituents. Thus ' $Fa \ \& \ Ga$ ' is true if and only if ' Fa ' is true and ' Ga ' is true. However, if we quantify, writing ' $(\exists x)(Fx \ \& \ Gx)$ ', then ' $\check{\text{Camp}}$,' no longer connects two sentences that have a definite truth-value. Some new interpretation for ' $\check{\text{Camp}}$,' must be found, so that we can understand how the semantic value of the unit ' $Fx \ \& \ Gx$ ' is determined.

Tarski ([1933](#)) found a systematic way to do this. Even though ' Fx ' has no truth-value, it has a truth-value *relative to an assignment of an individual i to ' x '*. ' $Fx \ \& \ Gx$ ' is true relative to an assignment if and only if ' Fx ' is true relative to that assignment and ' Gx ' is true relative to that assignment. We can also say that an assignment or an individual satisfies the open sentence ' $Fx \ \& \ Gx$ ' if and only if it satisfies ' Fx ' and satisfies ' Gx '. This solves the problem of quantifying into truth-functional contexts (see [Tarski's definition of truth](#)).

To interpret ' $\exists x \Box Fx$ ' we need to know when an individual i satisfies ' $\Box Fx$ '. It is not so immediately evident how to account for this, especially if you think that the necessity of ' Fa ' comes from features of ' Fa ' (for example, an analytic association of ' a ' with ' Fx ') rather than from features of the individual i that ' a ' denotes. An individual can be singled out by many different conditions, and perhaps some of those are analytically associated with ' Fx ' though others are not. For example, the condition ' $x = 3$ ' has ' $x > 7$ ' as a necessary consequence (that is, it is a necessary truth that whatever satisfies one satisfies the other); but ' x numbers the planets' does not have this necessary consequence. The condition ' $x > 7$ ' is necessarily connected with some ways of specifying nine but not with others. We can ask whether nine is necessarily greater than seven, but Quine finds no grounds for answering this question one way rather than another, because he believes that all necessity must be grounded in analytic consequence relationships and he finds no basis for favouring one specification of nine over the other. By this standard, either answer to the question of whether nine is necessarily greater than seven seems equally justified. Beginning with an account of necessity in terms of analyticity or logical truth should lead one to Quine's suspicions about quantification into this context.

The attempt to explain satisfaction of a modal condition in terms of analytic consequence seems doomed. But how else? Saying directly that some properties are necessary and others are not repels Quine. Quine sees such distinctions as an objectionable essentialism (see [Essentialism](#)), and accordingly he dismisses quantified modal logic.

Kripke provides an alternative account of what it is for an individual to satisfy ' $\Box Fx$ '. The account, in terms of possible worlds, has puzzled Quine; and puzzlement about existence of an individual in different possible worlds is closely related to Quine's puzzlement about essentialism. For existence in multiple possible worlds to have significance, it seems that there must be something that makes it possible for an individual to exist with some sets of properties and not others - some are essential and some are not. We must distinguish between properties that an individual has: for example, my cat satisfies ' x is an animal' and also ' x has eaten chicken livers', but she has the first property (being a cat) essentially and the second (having eaten chicken livers) only contingently. Kripke provides at least the formal framework for an account (1963), and in later work (1972) does much to defend the substantive metaphysics that goes with it.

Terence Parsons (1969) has pointed out that we can make at least one kind of distinction between essential and non-essential properties in a way that even Quine should not object to. Properties expressed by sentences of the form ' $\text{If } Fx, \text{ then } Fx$ ' will be essential to everything. Such essential properties seem unproblematic. We encounter problematic essentialism only if we find two individuals that differ in their essential properties. This point is limited in its scope, however. Within any fixed model verifying the assumption that all individuals have the same essential properties, any sentence with quantifying in is true in all of the same worlds as some sentence that lacks quantifying in. The resources that troubled Quine - quantification into modal contexts - seem to serve no expressive purpose (see [McKay 1975](#) and [Fine 1978b](#)).

4 Counterparts

Although Barney is eating a sandwich for lunch, he could have had a salad instead. Most modal theorists take that claim to involve the existence of a certain non-actual but possible situation: the possibility that Barney eats a salad for lunch.

David Lewis (1968) thinks that no such possible situation exists. Barney could have eaten a salad, but that really means that there is a possible world very much like this one in which the person most like Barney eats a salad. A non-actual world will never include Barney; it may, however, include a counterpart of Barney, an individual that is more like him than is any other in that world. The fact that one of his counterparts eats a salad makes it true that Barney could eat a salad, according to Lewis.

This approach has seemed unintuitive and ill-motivated to many (Plantinga (1974) and (Kripke 1972), for example). When considering what is possible for Barney, it seems natural to suppose that we are considering possibilities that involve Barney rather than someone else. In addition, it seems that in another possible world Barney could have been very different from the way he is while someone else was more like our actual Barney. As arguments against counterpart theory, however, such intuitive considerations are inconclusive. They might be answered by denials (citing a conflict of intuitions) or, for the last objection, by clarifications of the theory that elaborate the relevant similarity notion.

Formal considerations are also significant, however. The identity relation is transitive, but there is no reason to think that the counterpart relation is transitive, and this will produce different valid formulas. In the modal systems S4 and S5 (see [Modal logic](#)), for example, the following is ordinarily valid:

$$\forall x(\Diamond\Diamond Fx \rightarrow \Diamond Fx).$$

But as long as the counterpart relation is not transitive, we could not expect that to be valid, even in S5. (An individual i might have a counterpart that has a counterpart that satisfies ' Fx ' even though no counterpart of i satisfies ' Fx '.) Whether this difference would be an advantage or a drawback would need further consideration.

One technical feature is directly related to some of the intuitive considerations against counterpart theory. Lewis allows that an individual in one world might have two counterparts within a single world. If that is the case, then some laws involving identity will fail. Ordinarily, the principle of necessary identity is valid:

$$\Box \forall x \forall y (x = y \rightarrow \Box x = y).$$

And there seems to be a good argument for this, based on two other fundamental principles, the necessity of self-identity and Leibniz's Law (see **Identity**):

$$\Box \forall x \Box x = x \\ \forall x \forall y (x = y \rightarrow (\phi x \leftrightarrow \phi y))$$

One substitution instance of Leibniz's Law is this:

$$\forall x \forall y (x = y \rightarrow (\Box x = x \leftrightarrow \Box x = y)).$$

Given the assumption of the necessity of self-identity, the principle of necessary identity then follows. To defend counterpart theory, one must locate a fault with the necessity of self-identity or with this application of Leibniz's Law, or one must add features to the counterpart theory that will validate the principle of necessary identity (see Forbes 1985).

5 The logic of ability

Having surveyed some of the difficulties in interpreting modal logic, we now return to an illustration of the use of the possible worlds framework to shed light on inferences that are important for philosophy.

Although 'might' and 'can' are often thought of as possibility operators, they are clearly different. 'John might hit the target' and 'John can [is able to] hit the target' are very different, and Mark Brown's semantics for a logic of ability extends the possible worlds resources to make it clear what the difference is (1988).

Suppose that John is throwing darts, but does not reliably hit the target. Is John *able* to hit the target? He might hit it, but he does not have an ability that enables him to hit it reliably. In ordinary discussion, 'John is able to hit the target' is perhaps ambiguous. It could represent 'He might hit it', but is more likely to mean 'He has the ability to hit it reliably'. This suggests that there are two possibility operators, and these are distinguishable in Brown's system.

We can also give a more formal motivation for such a system. When we consider the context 'He has the ability to...reliably', we find that it does not obey the following law $K\Diamond$ of the most fundamental normal modal system K:

$$\Diamond(p \vee q) \rightarrow (\Diamond p \vee \Diamond q).$$

If John has the ability to hit the target reliably, then he is able to bring it about that he hits either the right half or the left half. It need not be true that he has the ability to hit the right half reliably, and it need not be true that he has the ability to hit the left half reliably. So $K\Diamond$ does not seem true, if ' $\Diamond$ ' is to represent an ability operator.

Also, in almost all standard modal systems (all that include T), we have this principle: $p \rightarrow \Diamond p$. But suppose that John will, by luck, hit the target. It does not follow that he has the ability to do that reliably. So $T\Diamond$ is not true, if ' $\Diamond$ ' is to represent an ability operator.

Brown introduces a family of new modal operators that we can add to the standard ones. We will use these symbols: $\Diamond$, $\Diamond$, $\Box$ and $\Box$. The semantics is based on a binary relation S , relating worlds in W to sets of worlds (that is, $S \subseteq (W \times \wp(W))$). We can define the new modal operators as follows (where we use ' K ' as a variable for sets of possible worlds):

$\Diamond p$ is true at w iff $\exists K$ such that wSK and $\forall w'$ such that $w' \in K$, p is true at w' . (Informally: John is able to bring it about that p reliably. John can perform some action such that p is sure to be true if that action is performed.)

$\Diamond p$ is true at w iff $\exists K$ such that wSK and $\exists w'$ such that $w' \in K$, p is true at w' . (John might bring it about that p . Some action that John can perform allows for the possibility of p .)

$\Box p$ is true at w iff $\forall K$ such that wSK , $\exists w'$ such that $w' \in K$, p is true at w' . (No matter what John does, p might be true.)

$\Box p$ is true at w iff $\forall K$ such that wSK , $\forall w'$ such that $w' \in K$, p is true at w' . (No matter what John does, p will

be true.)

Given this understanding of these operators, it is natural to make at least the following assumptions about the relation S , validating the formulas $\Diamond p \rightarrow \Diamond p$ and $p \rightarrow \Diamond p$:

$\forall w \forall K$ such that wSK , $K \neq \emptyset$. (Validates D $\Diamond$: $\Diamond p \rightarrow \Diamond p$)

(Nothing John can do annihilates all possibilities. Everything John can do has some possible outcomes.)

$\forall w \exists K$ such that wSK and $w \in K$. (Validates T $\Diamond$: $p \rightarrow \Diamond p$)

(The actual world is a possible outcome of at least one of John's actions.)

Focus first on $\Diamond$, the modal operator representing ability. If we think of an action as being represented by a set of possible outcomes, then we can consider each set K of possible worlds such that wSK to correspond to the outcomes of some action that the agent under consideration might take. Each such set K would be a set of possible worlds that might be actual if the agent performed some particular action. So if the agent can hit the dartboard, but has no control over where, then a set of possible worlds that vary in where the agent hits the dartboard would be a set K such that wSK , and it could correspond to the action of throwing a dart. To say that $\Diamond A$ is to say that there is some action (that is, some set K of outcomes; that is, some set K such that wSK) that guarantees A (that is, A is true throughout K). To say that the agent can hit the dartboard is to say that there is some such action (that is, some set K such that wSK , and throughout K the agent hits the dartboard).

The word 'could' is associated with possibility and ability. Three of our new modal operators involve possibility, and we can give an example that is three ways ambiguous:

He could win the lottery.

This could mean that he bought a ticket, and so nothing he can do would guarantee to make him a non-winner, that is, $\Box A$. This could mean that he can buy a ticket, and then he might win, that is, $\Diamond A$. Or this could mean that he has special powers such that if he exercised them he would win, that is, $\Diamond A$. (The King could win the lottery (if he wanted to), but he hates to interfere in the peasants' fun.)

Thus we have a resource for disambiguation and for the exploration of logical relationships that are central to some philosophical work in ethics and metaphysics. Even though we have seen how many puzzles arise in the acceptance of the possible worlds framework, we should not lose sight of the many reasons we have to value it.

See also: [Logical and mathematical terms, glossary of](#); [Modal operators](#); [Montague, R.M.](#)

THOMAS J. MCKAY

References and further reading

- Blackburn, S.** (1993) 'Morals and Modals', in *Essays in Quasi-Realism*, Oxford: Oxford University Press. (Discusses the problem of §1: what makes modal claims true.)
- Brown, M.** (1988) 'On the Logic of Ability', *Journal of Philosophical Logic* 17: 1-26. (An extension of standard modal logic to accommodate statements of ability, such as 'John can [is reliably able to] hit the target', and related concepts. Discussed in §5.)
- Fine, K.** (1978a) 'Model Theory for Modal Logic. Part I: The *De Re/De Dicto* Distinction', *Journal of Philosophical Logic* 7: 125-56. (Relates technical results to some philosophical considerations.)
- Fine, K.** (1978b) 'Model Theory for Modal Logic. Part II: The Elimination of the *De Re*', *Journal of Philosophical Logic* 7: 277-306. (Technical results and philosophical issues related to those discussed in §3.)
- Fine, K.** (1981) 'Model Theory for Modal Logic. Part III: Existence and Predication', *Journal of Philosophical Logic* 10: 293-307. (Technical results and philosophical issues related to a version of actualism, a view that ascribes a special status to actual or existent objects. Related to issues discussed in §2.)
- Forbes, G.** (1985) *The Metaphysics of Modality*, Oxford: Oxford University Press. (A discussion of metaphysical issues associated with modal logic.)
- Kaplan, D.** (1989) 'Demonstratives', in *Themes from Kaplan*, ed. J. Almog, J. Perry and H.K. Wettstein, Oxford: Oxford University Press. (A classic discussion of demonstrative reference, with a semantics that builds on possible worlds models. Makes clear the distinction between necessity and analyticity mentioned in connection

with §3.)

- Kripke, S.A.** (1963) 'Semantical Considerations on Modal Logic', *Acta Philosophica Fennica* 16: 83-94; repr. in L. Linsky (ed.) *Reference and Modality*, Oxford: Oxford University Press, 1971. (This presents the semantic framework for modal logic.)
- Kripke, S.A.** (1972) 'Naming and Necessity', in D. Davidson and G. Harman (eds) *Semantics of Natural Language*, Dordrecht: Reidel, 252-355. (This important article illuminates the concept of metaphysical necessity and identifies important principles of identity. There is material relevant to every section of this entry.)
- Lewis, D.K.** (1968) 'Counterpart Theory and Quantified Modal Logic', *Journal of Philosophy* 65: 113-26. (Presents and defends counterpart theory. See §4.)
- Lewis, D.K.** (1973) *Counterfactuals*, Cambridge, MA: Harvard University Press. (In discussion of the main topic, conditionals, there is a valuable development of the possible worlds framework. There is also some general discussion of the nature of possible worlds and of counterpart relations.)
- Lewis, D.K.** (1986) *On the Plurality of Worlds*, Oxford: Blackwell. (Further development of Lewis' view of possible worlds, with some important challenges to other views of modality.)
- Linsky, L. (ed.)** (1971) *Reference and Modality*, Oxford: Oxford University Press. (A collection of important articles on the topics of §§2, 3.)
- McKay, T.** (1975) 'Essentialism in Quantified Modal Logic', *Journal of Philosophical Logic* 4: 423-38. (Results mentioned in §3, concerning Quine's linking of modal logic and essentialism.)
- Montague, R.** (1974) *Formal Philosophy*, New Haven, CT: Yale University Press. (In these papers, Montague developed the semantic framework of possible worlds in beginning to give a more general semantics for natural language.)
- Parsons, T.** (1969) 'Essentialism and Quantified Modal Logic', *Philosophical Review* 78: 35-52; repr. in L. Linsky (ed.) *Reference and Modality*, Oxford: Oxford University Press, 1971, 73-87. (Anti-essentialist models introduced as part of a response to Quine. See §3.)
- Plantinga, A.** (1974) *The Nature of Necessity*, Oxford: Oxford University Press. (An excellent introduction to topics in the metaphysics of modality. Relevant to §§1-4 of this entry.)
- Quine, W.V.** (1961) 'Reference and Modality', in *From a Logical Point of View: Nine Logico-Philosophical Essays*, Cambridge, MA: Harvard University Press, 2nd edn, 1980; repr. in L. Linsky (ed.) *Reference and Modality*, Oxford: Oxford University Press, 1971. (Misgivings about the interaction of modality with reference and quantification. Discussed in §3.)
- Quine, W.V.** (1966) 'Three Grades of Modal Involvement', in *The Ways of Paradox*, New York: Random House. (Misgivings about the interaction of modal sentence operators with reference and quantification. Discussed in §3.)
- Stalnaker, R.** (1968) 'A Theory of Conditionals', in N. Rescher (ed.) *Studies in Logical Theory*, Oxford: Blackwell. (A very readable presentation of a theory of conditionals that builds on the semantic framework of possible worlds.)
- Tarski, A.** (1933) *Pojęcie prawdy w językach nauk dedukcyjnych*, Warsaw; trans. J.H. Woodger (1956) 'On the Concept of Truth in Formalized Languages', in *Logic, Semantics, Metamathematics*, ed. J. Corcoran, Indianapolis, IN: Hackett Publishing Company, 2nd edn, 1983, 152-278. (The semantics for quantification for non-modal logic. See §3.)

Modal operators

Modal logic is principally concerned with the alethic modalities of necessity and possibility, although this branch of logic is applied to a wide range of linguistic and conceptual phenomena, including natural language semantics, proof theory, theoretical computer science and the formal characterization of knowledge and belief. This wide range of application stems from the basic form of modal assertions, such as ‘it is necessarily the case that ϕ ’, where an entire statement ϕ is embedded within a context possessing rich logical structure.

When constructing a formal representation of these embedding contexts, there are several choices concerning their specific symbolic form. The most standard approach symbolizes modal contexts as operators, which combine directly with formulas of the object language to yield new formulas. The primary alternative to this approach is to treat modal contexts as predicates, which attach not to formulas directly, but to names of formulas, and thereby attribute a metalinguistic property to a syntactic object. A variation on the operator approach, which assumes the interpretive framework of possible worlds semantics, is to treat modal contexts as quantifications over possible worlds. Finally, a variation on the predicate approach is to analyse modal contexts as predicates of propositions rather than as predicates of syntactic objects.

1 Monadic connectives

On the most standard approach, the modal operators ‘ $\Box$ ’ for necessity and ‘ $\Diamond$ ’ for possibility are introduced as new logical constants (although only one need be taken as primitive). These modal operators are syncategorematic expressions, since they are not assigned an explicit reference, rather their semantic import is carried in a recursive clause of the formal definition of truth. In terms of the syntactic formation rules, the modal operators behave as monadic connectives, structurally comparable to negation in classical logic. The modal assertion ‘it is necessarily the case that ϕ ’ will have the symbolic form ‘ $\Box\phi$ ’, where the necessity connective ‘operates’ on the formula ϕ to yield the new formula ‘ $\Box\phi$ ’. However, unlike the negation connective, the modal operators are not truth functional: it is not possible to determine the truth-value of a modal statement simply on the basis of the truth-value of the formula it embeds. In this sense, introduction of the modal operators constitutes a significant departure from the pure extensionality of classical logic, and this has led to criticism of their legitimacy, most notably from W.V. Quine.

Quine has argued that non-extensional contexts are logically permissible only if construed as ‘referentially opaque’, that is, the constituent symbols within these contexts are treated as occurring non-referentially. Quotation is a paradigmatic case of such a context since, for example, in the true statement “‘John walks’ is a short sentence”, the occurrence of the string ‘John’ does not refer to anyone (see [Use/mention distinction and quotation](#)). Quine (1953) argues that since modal contexts are non-extensional, they are properly viewed as quotational, wherein the embedded sentence ϕ is mentioned rather than used (this theme will be continued in §3; see [Propositional attitudes §2](#)). However, treatment of modal contexts as operators suggests that the logical structure of the embedded formula is semantically relevant, even though it occurs within a non-extensional context. Indeed, since ‘ $\Box$ ’ operates at the same linguistic level as the formula ϕ , it follows that the semantic value of ϕ should directly contribute to the compositional value of ‘ $\Box\phi$ ’. Thus the ‘ $\Box$ ’ operator must utilize information not contained in the extension of the embedded expression.

2 Sorted quantifiers

According to possible worlds semantics for modal operators, introduced by Saul Kripke (1959), the non-extensional value of modal statements is determined by appeal to the extension of the embedded expression in alternative possible worlds (see [Semantics, possible worlds](#)). The sentence ‘ $\Box\phi$ ’ is evaluated as true at a given world just in case ϕ is true in all worlds possible relative to the given world, while the sentence ‘ $\Diamond\phi$ ’ is true just in case there is at least one possible world in which ϕ is true. So in this standard semantic framework, the truth definition for modal statements is specified in the metalanguage by quantifying over the set of possible worlds; the truth of an assertion of necessity is expressed in terms of a universal quantification, while an assertion of possibility is evaluated as an existential quantification. Somewhat more formally, a Kripke model M comes equipped with a set W of worlds (or points or indices) and a binary relation R on W such that a world $w' \in W$ is possible relative to a world $w \in W$ just in case wRw' . In the simplest case, where R is universal (that is, every

world in W is possible relative to itself and every other), reference to R can be dropped. Then ' $\Box\phi$ ' is true in w just in case ϕ is true in all worlds $w' \in W$, and ' $\Diamond\phi$ ' is true in w just in case there is at least one $w' \in W$ in which ϕ is true.

It is therefore possible to recast the logical form of modal statements as quantifiers rather than operators, where the structure of the truth definition is explicitly manifested in the structure of the object-language formula. In the general case, this will require the use of many-sorted quantificational logic, where the domain of discourse is divided into distinct subdomains, and typographically different variables are used to quantify over these respective subsets. In the modal case, the set W of possible worlds will form one such subdomain, and a distinct type of intensional variable will range over this set. Thus where the extension of an expression ψ varies according to the index of evaluation, this same 'intensional' effect can be mirrored by adding a variable to yield the expression $\psi(i)$, where i ranges over possible worlds. The interpretation function will assign to ψ a value which gives the standard type of extension when applied to worlds as arguments. To translate operator modality into the quantifier formulation, a straightforward mapping is defined where $(\psi)^* = \psi(i)$ for nonlogical constants ψ , where logical form is preserved for the extensional connectives and where $(\Box\phi)^* = \forall i(\phi)^*$ and $(\Diamond\phi)^* = \exists i(\phi)^*$. In this manner, the necessity operator is explicitly recast as a universal quantifier, and possibility is represented as an existential quantifier. The quantifier approach is perhaps most attractive as an alternative to intensional type theory (see Montague, R.M.), where the different sorts of variables correspond to distinct logical types.

3 Predicates of expressions

While the quantifier interpretation is a natural variation on the use of operators (given the framework of possible worlds semantics), the predicate approach to modality constitutes a more divergent analysis. Under this development, modal contexts are symbolized by a monadic predicate, say N , which is categorematic and receives a normal first-order extension (that is, a set of objects). But admissible extensions of N must satisfy the constraints imposed by its intended meaning as a modal device, and in this sense N is comparable to the binary predicate symbol 'Čequals;' normally used to designate the identity relation.

On the predicate approach to modality, syntactic expressions are themselves named by singular terms, and the necessity predicate N attaches to such terms to yield new atomic formulas. Quotation is often adopted as a convenient method for constructing terms that name expressions. Thus if ' $\ulcorner\phi\urcorner$ ' is the official quotation term denoting the object-language formula ϕ , then the assertion that ϕ is necessary will have the form $N(\ulcorner\phi\urcorner)$. In this manner, modal formulas are viewed as attributing a metalinguistic property to a linguistic expression treated as an object of discourse. The two most natural properties to attribute in this context are that the object-language formula is valid with respect to some class of models, or that it is provable in some formal system. On this latter reading, the predicate approach is quite suitable for applications to proof theory (see Gödel's theorems §6) and for applications to the symbol-processing accounts of epistemic contexts common in artificial intelligence.

On the predicate analysis, the named formula is mentioned rather than used, and this is amenable to the Carnap-Quine position that the logic of modalities should be expressed in an extensional metalanguage (see Carnap 1937), and that the attempt to express modality in an intensional object language rests on a confusion between use and mention. It is also compatible with Quine's view (see §1) that modal contexts are referentially opaque. The fact that the modal predicate attaches to a term denoting a formula indicates that there are no free variables inside a modal context, and this would prohibit the articulation of quantified modal logic. Therefore, Quine has argued that the only legitimate development of 'intensional' logic is comprised by that fragment which can be rendered in terms of the extensional clarity of a metalinguistic predicate.

However, Montague (1963) has subsequently demonstrated that for any language endowed with the 'self-referential' capacities of elementary number theory, the addition of even very weak modal assumptions leads to inconsistency under the predicate approach. In systems that extend the theory Q of Robinson arithmetic, Gödel's diagonal lemma establishes that for any open formula $\beta(x)$ with one free variable, there exists a closed formula θ such that $\vdash \theta \leftrightarrow \beta([\theta])$, where $[\theta]$ is the Gödel numeral code for θ . If necessity is formalized as a predicate $N(x)$ within such a system, then the diagonal lemma establishes that there will be a sentence μ such that $\vdash \mu \leftrightarrow \neg N([\mu])$, where μ 'says of itself' that it is not necessary. μ is the modal analogue of the celebrated Gödel sentence that asserts its own unprovability, and of the liar sentence that underpins Tarski's theorem on the undefinability of truth (see Semantic paradoxes and theories of truth). And in a manner analogous to the

Gödel-Tarski results, μ can be used to deduce a contradiction, given fairly minimal principles governing the logic of N .

In light of these negative results, the predicate view of modality must be approached with some care. A crucial point is that not all the formulas in which N occurs will correspond to statements that can be articulated using the resources of the operator. In particular, there is no operator counterpart to the ‘self-referential’ sentence μ , since the ‘ $\square$ ’ connective must apply to formulas rather than to their names. And if radically new formulas such as μ are allowed to instantiate the relevant modal schemata, then new axioms are added to the modal system, and this in turn yields an unwanted increase in deductive power. Thus it is possible to avoid inconsistency by inverting the Quinean perspective and defining the predicate system so that it precisely mirrors the expressive-deductive power of the operator.

4 Predicates of propositions

In addition to treating them as predicates of syntactic objects, it is also possible to formalize modal contexts as predicates of semantic objects. In systems of higher-order intensional logic such as Montague’s system IL, semantic value is divided along Fregean lines into both an extensional and an intensional component (see [Sense and reference](#)). Montague’s scheme utilizes Carnap’s insight that the intensional component of semantic value can be modelled as a function from possible worlds to extensions. The intension of a sentence is a proposition, while its extension is a truth-value, and hence a proposition is modelled as a (characteristic function of a) set of possible worlds, namely that set of worlds in which the sentence is true. The object language is equipped with an intensional operator ‘ $\hat{}$ ’ which, when attached to an expression, yields a term that denotes the intension of the expression. So the term $\hat{\phi}$ denotes the proposition expressed by ϕ , which is the set of possible worlds in which ϕ is true. Necessity can then be rendered as a higher-order predicate *Nec*, which applies to sets of worlds and is meant to assert that the set in question is identical to the set of all worlds in the model. In this framework, the statement that ϕ is necessary will then have the form *Nec*($\hat{\phi}$), which asserts that the set of worlds constituting the intension of ϕ is identical to the set of all worlds in the model, and this is equivalent to the standard operator interpretation that ϕ is true in all possible worlds.

See also: [Logical and mathematical terms, glossary of](#); [Modal logic](#); [Modal logic, philosophical issues in](#); [Possible worlds §2](#)

PAUL SCHWEIZER

References and further reading

- Carnap, R.** (1937) *The Logical Syntax of Language*, London: Routledge & Kegan Paul. (Expounds the view that modality should be expressed in an extensional metalanguage.)
- Dowty, P., Wall, R. and Peters, S.** (1981) *Introduction to Montague Semantics*, Dordrecht: Reidel. (Provides an introduction to Montague’s system of intensional type theory.)
- Gallin, D.** (1975) *Intensional and Higher-Order Modal Logic*, Amsterdam: North Holland. (Gives a treatment of modal operators as quantifiers within the framework of type theory.)
- Kripke, S.A.** (1959) ‘A Completeness Theorem in Modal Logic’, *Journal of Symbolic Logic* 24: 1-15. (Introduces possible worlds semantics for modal logic.)
- Montague, R.** (1963) ‘Syntactical Treatments of Modality, with Corollaries on Reflexion Principles and Finite Axiomatizability’, in R. Thomason (ed.) *Formal Philosophy*, New Haven, CT: Yale University Press, 1974. (Presents inconsistency proofs for predicate version of modal logic.)
- Quine, W.V.** (1953) ‘Three Grades of Modal Involvement’, in *The Ways of Paradox and Other Essays*, Cambridge, MA: Harvard University Press, 1966. (Advocates the view that modal contexts are quotational.)
- Schweizer, P.** (1992) ‘A Syntactical Approach to Modality’, *Journal of Philosophical Logic* 21: 1-31. (Addresses Montague’s argument that the predicate approach to modality is inconsistent.)
- Skyrms, B.** (1978) ‘An Immaculate Conception of Modality’, *Journal of Philosophy* 75: 77-96. (Addresses Montague’s argument that the predicate approach to modality is inconsistent.)

Model theory

Model theory studies the relations between sentences of a formal language and the interpretations (or ‘structures’) which make these sentences true or false. It offers precise definitions of truth, logical truth and consequence, meanings and modalities. These definitions and their consequences have revolutionized the teaching of elementary logic.

Model theory also forms a branch of mathematics concerned with the ways in which mathematical structures can be classified. This technical work has led to philosophically interesting results in at least two areas: it has thrown light on the nature of the set-theoretic universe, and in nonstandard analysis it has suggested new forms of argument (where we prove something different from what we intended, but then use a general model-theoretic argument to change the result into what we wanted).

The word ‘model’ has many other uses. For example, model theory is not about scientific theories as models of the world. It is also a controversial question - not considered here - how model theory is connected with the ‘mental models’ which appear in the psychology of reasoning.

1 Basic ideas

In 1954 Alfred Tarski announced that a new branch of metamathematics had come into being. He called it ‘model theory’; it studied ‘mutual relations between sentences of formalized theories and mathematical [structures] in which these sentences hold’. Model theory quickly became an active area of research on the boundary between logic and other branches of mathematics (see §5 below). The name ‘model theory’ is sometimes applied more broadly, particularly by philosophers, to include any inquiry which uses formal languages, structures and a truth definition. For example, a model-theoretic approach to English semantics is one which treats English as a formal language and attaches meanings to English phrases in terms of possible worlds, regarding these worlds as mathematical structures.

When Tarski spoke of a ‘sentence of a formalized theory’, he meant what logicians have also called a ‘sentence schema’, that is, a string of meaningful words and meaningless symbols (sometimes called schematic variables), which becomes a meaningful sentence when the symbols are given meanings. For example, the string

(1) Everything which is a P is a Q

is a sentence schema. If we interpret ‘ P ’ as meaning ‘cow’ and ‘ Q ’ as meaning ‘quadruped’, then the schema comes to mean the same as the sentence ‘Every cow is a quadruped’. In model theory we always assume that the word ‘everything’ needs interpretation too; the interpretation will say what set of things counts as everything, and any set may be allowed. For example, if we take ‘everything’ to range over the set of objects in China, then the original string comes to mean ‘Every object in China which is a cow is a quadruped’.

A ‘mathematical structure’ is a way of interpreting the symbols in a schema. It consists of a set of objects which is the range of quantifier words such as ‘everything’ and ‘there exists’; and a family of sets, relations, functions and so on which express the meanings of the schematic variables. The set of objects is called the ‘domain’ or ‘universe’ of the structure, and its members are called the ‘elements’ of the structure. For example, the structure which gives our interpretation of (1) has the set of objects in China for its domain; it attaches the set of cows in China to the symbol ‘ P ’ and the set of quadrupeds in China to the symbol ‘ Q ’. (More formally, we can regard a structure M as an ordered pair $\langle \text{dom}(M), Y \rangle$ where $\text{dom}(M)$ is the domain of M and Y is a function which takes each interpreted symbol s to its interpretation $Y(s)$, also written s^M . Many mathematical objects can also be thought of as structures in this sense; thus a group G is a pair $\langle X, Y \rangle$ where X is the set of elements of G and Y is a function taking the group product symbol ‘ $\cdot$ ’ to the group multiplication ‘ $\cdot^G$ ’.)

In our schema (1), the meaningful expressions are those written in English. Model theorists often use logical symbols for the meaningful words. For example, ‘and’ is often written ‘ $\wedge$ ’ or ‘ $\&$ ’; for ‘it is not true that’ one writes ‘ $\sim$ ’ or ‘ $\neg$ ’; one writes ‘ (x) ’ or ‘ $\forall x$ ’ to mean ‘for every element x ’. When this is done, a sentence schema is the same thing as a sentence of formal logic with uninterpreted non-logical symbols (see [Predicate calculus](#)).

2 Truth and satisfaction

The central device of model theory is the truth relation, written ‘ $\models$ ’. (Pronunciations vary, but ‘double turnstile’ is safe, using ‘turnstile’ for ‘ $\vdash$ ’.) Suppose ϕ is a sentence schema and M is a structure which interprets ϕ . Interpreted, ϕ says something which is either true or false. If it is true, we say that ϕ is true in M , or that M is a model of ϕ ; in symbols, $M \models \phi$. Tarski showed how one can give a precise mathematical definition of the relation ‘ $\models$ ’. His definition is correct in the sense that it agrees with our intuitive notions about when an interpretation M makes a formula ϕ into a true sentence. (See **Tarski’s definition of truth** for the context of Tarski’s definition. See also **Semantics, game-theoretic** for an interesting variant.)

In defining ‘ $\models$ ’, Tarski assumed that each structure M is a set-theoretic object and that the schemas ϕ are sentences of a conventional formal logic. Also M must interpret every non-logical symbol in ϕ . It is usual to ensure this by stating Tarski’s definition relative to a set S of non-logical symbols; we assume that M interprets just these symbols, and that ϕ uses no non-logical symbols besides those in S . This set S is then called the ‘signature’ (or similarity type) of M and ϕ .

In fact Tarski defined a slightly more complicated relation, namely satisfaction. Suppose M is a structure, $a_1, \dots, a_n$ are elements of M and $\phi(x_1, \dots, x_n)$ is a formula with free variables $x_1, \dots, x_n$. We write $M \models \phi[a_1, \dots, a_n]$ (pronounced ‘ a_1 to a_n satisfy ϕ in M ’) to mean that when M is used to interpret the symbols of ϕ apart from the free variables, and x_i is read as the name of the element a_i for $1 \leq i \leq n$, then ϕ becomes a true sentence. The truth relation is the special case where $n = 0$. (The set of all n -tuples $(a_1, \dots, a_n)$ such that $M \models \phi[a_1, \dots, a_n]$ holds is called the relation (or set) defined in M by ϕ .)

Tarski’s definition of satisfaction has several clauses, depending on the form of the formula ϕ . For example, if ϕ is of the form $\psi \wedge \chi$ then $M \models \phi[a_1, \dots, a_n]$ holds if and only if $M \models \psi[a_1, \dots, a_n]$ and $M \models \chi[a_1, \dots, a_n]$ both hold. If ϕ is of the form $\forall x_2 \psi(x_1, x_2)$, then $M \models \phi[a_1]$ holds if and only if for every element b of M , $M \models \psi[a_1, b]$ holds. These and similar clauses allow us to break ϕ down into its smallest component formulas, known as atomic formulas. For atomic formulas, Tarski’s definition refers directly to the interpretations of symbols given by the structure M .

By applying Tarski’s definition, we can reduce the statement that $M \models \phi[a_1, \dots, a_n]$ to a purely set-theoretic statement. To illustrate this, let $\phi(x)$ be the first-order formula ‘ $\forall y(y = c \wedge Py \rightarrow y = x)$ ’, and let M be a structure with a set P^M to interpret the symbol ‘ P ’ and an element c^M to interpret the constant c . Suppose b is an element of M . Then according to Tarski’s definition of satisfaction (with a very small simplification),

$$(2) \quad M \models \phi[b]$$

holds if and only if

$$(3) \quad \{c^M\} \cap P^M \subseteq \{b\}.$$

(Intuitively, every member of the set P^M which is equal to c^M is equal to b . The schema ϕ defines in M the set $\{c^M\}$ if c^M is in P^M , and the set of all elements of M otherwise.)

3 Classes defined by axioms

Model theory grew out of a development in mathematics. Around the turn of the twentieth century it was becoming common to define classes of structures in terms of the axioms which are true in them; such classes are said to be axiomatically defined. The classes of groups, rings and Boolean algebras are three examples (see **Boolean algebra**). Certain questions arise naturally in this context. Let us note three. First, what can we tell about an axiomatically defined class by looking at the syntactic form of its axioms? Second, given an axiomatically defined class, when is one of the symbols redundant in the sense that its meaning in any structure in the class is determined by the meanings of the other symbols? Third, given a set of axioms T , one of which is ϕ , when is ϕ redundant in the sense that the class axiomatically defined by T would not be altered if we left out ϕ ? These are all typical questions of model theory.

By a theory we mean a set of sentences of a formal language. We say that a structure M is a model of the theory T if M is a model of every sentence in T . So the class of all models of T is the axiomatically defined class whose set

of axioms is T . First-order model theory, which is by far the largest part of the subject, studies the case where T is a set of first-order sentences. Many interesting mathematical classes of structures - including groups, rings and Boolean algebras - are defined by first-order axioms.

Here are two sample theorems of first-order model theory, which illustrate the first and second questions above. We shall turn to the third question in the next section.

Two structures are said to be ‘elementarily equivalent’ if they are models of exactly the same first-order sentences. Our first theorem states that if M is a structure with infinitely many elements and λ is an infinite cardinal, then we can cut down M (if it has more than λ elements) or add new elements (if it has fewer) so as to form a structure N which is elementarily equivalent to M and has exactly λ elements. This theorem is known as the downward Löwenheim-Skolem theorem (when elements are removed) or the upward Löwenheim-Skolem theorem (when elements are added). The upward and downward theorems together are equivalent to the axiom of choice, which is an indication that in general it is impossible to give an explicit description of N . The upward theorem follows from the compactness theorem of §4.

If T is a first-order theory, and a symbol R of the signature of T is redundant in the sense described at the start of this section, then we say that R is implicitly definable relative to T . Our second theorem states that if R is implicitly definable relative to T , then it is explicitly definable in the sense that there is some first-order sentence which is true in all models of T and which says, ‘ R means such-and-such’. (See [Beth’s theorem](#) and [Craig’s theorem](#) for further details.) In practice this theorem is generally used backwards: we are given that R is not explicitly definable relative to T , and we use the theorem to find two models of T which agree in the interpretations of all symbols except R .

4 Semantic and syntactic consequence in first-order logic

The question of redundant axioms goes back several hundred years. Already in the thirteenth century Persian geometers were trying to show Euclid’s parallel postulate to be redundant in his axioms for geometry. In fact it is not, and proofs of this fact by Beltrami and others in the late nineteenth century gave a boost to the development of model theory. Beltrami introduced the word ‘interpretation’ in 1868.

Some notation will help. If T is a theory and ϕ a sentence, we write

$$(4) \quad T \models \phi$$

to mean that every model of T is also a model of ϕ . (4) is sometimes read ‘ ϕ is a semantic consequence of T ’. (This double use of ‘ $\models$ ’ is unfortunate, but the usage has become firmly entrenched. In (4) there is a theory to the left of ‘ $\models$ ’, not a structure as in (2).) So ϕ is redundant in the theory consisting of T and ϕ , if and only if $T \models \phi$.

The central theorem here is Gödel’s completeness theorem, proved in his doctoral dissertation of 1929 and published in 1930. It states that if T and ϕ are first-order, then (4) holds if and only if ϕ is deducible from T in a standard proof calculus for first-order logic. We write $T \vdash \phi$ when this condition holds; so Gödel’s theorem says that the structural relation ‘ $\models$ ’ (in the sense of (4)) is equivalent to the purely syntactic relation ‘ $\vdash$ ’. Like all later logicians, Gödel proved his theorem by assuming that ϕ is not deducible from T and constructing a model of T in which ϕ is false. Later logicians have usually built the model out of the expressions of the first-order language; Gödel, perhaps influenced by Hilbert’s constructions of counterexamples in geometry, used numbers as the elements of the model and showed that appropriate relations could be found to interpret the symbols of the language. (See [Gödel’s theorems](#). Gödel proved his theorem for one particular proof calculus and a countable signature, but both these restrictions are unnecessary.)

Of the many applications of Gödel’s completeness theorem, two are worth noting here. First, when T is the empty theory in (4), we leave it out and write simply ‘ $\models \phi$ ’; and likewise with ‘ $\vdash$ ’. Compare the forms of the following two statements.

$$(5) \quad \vdash \phi, \text{ that is, there is a proof of } \phi$$

$$(6) \quad \models \phi, \text{ that is, there is not a model of } \neg\phi.$$

By Gödel’s theorem, (5) and (6) are equivalent, both saying that ϕ is a theorem (or valid sentence) of first-order

logic. The difference between the forms of (5) and (6) is very convenient. If ϕ is a theorem, according to (5) we can show it by giving a proof of ϕ . If ϕ is not a theorem, then by (6) we can show it by giving a structure in which ϕ is not true. Either way, we know what we have to produce in order to establish the fact. (A strengthening of the completeness theorem, proved independently by Hasenjäger and Kleene in 1952, allows us to do better. If ϕ is a first-order sentence but not a theorem, then there is a model of $\neg\phi$ which consists of the natural numbers together with some arithmetically definable sets, relations and functions. So in the second case we can specify the model by giving some arithmetical formulas.)

Now that we know Gödel's theorem, we can try to prove a first-order sentence ϕ by showing that its negation $\neg\phi$ has no model. This idea leads quickly to tableau proofs (see [Natural deduction, tableau and sequent systems §4](#)). Of course the tableau method itself is a purely syntactic calculus. But it can be motivated very directly in terms of the model-theoretic relation (4); with a good class of students one can explain the principle and leave it as a homework exercise to work out the rules.

Second, suppose T is a first-order theory and every finite subset of T has a model. Then there is no finite subset T' of T such that every model of T' is a model of the contradictory statement ' $\neg\forall x x = x$ ' (since no structure is a model of this statement). So by Gödel's theorem, T has no finite subset from which we can formally prove ' $\neg\forall x x = x$ '. But then we cannot prove ' $\neg\forall x x = x$ ' from T at all, since any formal proof would use only a finite part of T . Hence by Gödel's theorem again, it is not true that every model of T is a model of the contradiction; in other words, T has a model. By this roundabout route we have deduced the 'compactness theorem': if every finite subset of a first-order theory has a model, then the whole theory has a model. This is the single most important theorem of model theory. All the applications in the next section use it.

5 First-order model theory: four applications

This section relates some of the mathematical successes of first-order model theory. It can be skipped without breaking continuity.

First, suppose T is a first-order theory, and for every natural number n there is a model M_n of T in which there are at least n elements which satisfy the formula Px . For each n we can write a first-order sentence which says, 'There are at least n elements x such that Px '; let U be the set of these sentences for all n . Our assumption about T implies that if we take just finitely many sentences from U , then there is some model of T in which all these sentences are true. Then the compactness theorem tells us that there is a model of T in which all the sentences are true, so that infinitely many elements satisfy Px . This argument has many applications. It shows, for example, that if T has arbitrarily large finite models then T has an infinite model. It also shows that there is a structure N which is elementarily equivalent to the field R of real numbers (and in fact contains it), but which has infinitesimal elements; these are elements greater than 0 but less than 1, $\frac{1}{2}$, $\frac{1}{3}$ and so on. Since N contains infinitesimals, we can do calculus in N in the way that Leibniz intended, calculating dy/dx as a ratio of two infinitesimal numbers dy and dx (this is a slight oversimplification). Suppose that we can use infinitesimals to prove that a certain first-order sentence ϕ is true in N . Then ϕ must be true in R too, since R and N are elementarily equivalent. This ingenious way of doing calculus has sprouted into a new branch of mathematics (see [Analysis, nonstandard](#)).

In 1982 Shelah gave a very different application of model theory. Confining himself to structures of countable signature, he looked at elementary equivalence classes, that is, classes that consist of a structure and every other structure elementarily equivalent to it. Two structures are said to be isomorphic if one is an exact copy of the other. (Formally, M is isomorphic to N if there is a bijection from the domain of M to that of N , which takes the interpretation s^M of each symbol s of the signature to the interpretation s^N .) Shelah proved that a dichotomy holds: each elementary equivalence class is either very good or very bad. 'Very good' means that we can classify all the structures in the class, up to isomorphism, by means of a few numerical invariants (such as dimension in the case of vector spaces). 'Very bad' means that we can construct two structures in the class which are not isomorphic but are extremely hard to distinguish from each other. Shelah sees this work as a step towards bringing the more chaotic areas of mathematics under control.

In 1992 Hrushovski and Zil'ber proved a theorem which says that algebraically closed fields can be recognized through certain model-theoretic properties of their Zariski topologies. Hrushovski then used this result to prove an open case of the geometric Mordell-Lang conjecture (an important conjecture in diophantine geometry). Zil'ber

had long been convinced that there must be some kind of characterization of algebraically closed fields in terms of their model-theoretic properties, partly because these fields are such fundamental objects in mathematics. It was also partly because Macintyre had already shown that if M is an infinite field, then M is algebraically closed if and only if the set T of all first-order sentences true in M is λ -categorical for every uncountable cardinal λ - that is, that any two models of T with exactly λ elements must be isomorphic. (A theory is said to be categorical if all its models are isomorphic. By the upward Löwenheim-Skolem theorem, if a first-order theory has a model with infinitely many elements, then it cannot be categorical, though it may be λ -categorical for some cardinal λ .)

Finally, one achievement of the twentieth century has been to base all mathematics on axiomatic set theory. A theorem is reckoned to be proved if we know in principle how to write it as a formal deduction from the axioms of ZF (Zermelo-Fraenkel set theory). Unfortunately ZF seems to leave unanswered many natural questions about sets and infinite cardinals. Since the 1960s, set theorists have studied these questions by examining models of the ZF axioms. For example, in 1963 Paul Cohen showed that the axiom of choice is not deducible from the other axioms of ZF, by constructing models of these other axioms in which the axiom of choice is false. Model theorists have devised several techniques for constructing interesting models of a given theory, and two of these techniques have led to other advances in set theory. Ultraproducts, which are good for constructing ‘fat’ models in which there are as many types of element as possible, are an essential tool for studying the colossal (but possibly non-existent) cardinals that set theorists call large cardinals. The technique of indiscernibles, which constructs ‘thin’ models with as few types of element as possible, is used to show the enormous differences between Gödel’s constructible universe and a universe of sets which contains very large cardinals. (See [Set theory](#).)

6 Limitations of first-order model theory in mathematics

First-order logic has a delicate balance of expressiveness: it can say things about a fixed number of elements at a time, but it is very bad at expressing statements which involve infinitely many, or arbitrarily large finite numbers of, elements. For example, it is easy but boring to write down a first-order sentence which says, ‘The number of elements x such that Px is exactly 10,556,001’. By contrast there is no first-order sentence that says, ‘There are infinitesimals’. (The definition of infinitesimals involves the infinitely many elements 1 , $\frac{1}{2}$, $\frac{1}{3}$ and so on.) Also we saw that no first-order sentence expresses ‘There are just finitely many x such that Px ’. Even if we restrict attention to finite structures, there is no first-order sentence which is true in just those structures for which the set of elements satisfying the formula ‘ Px ’ has an even number of elements.

We have other formal languages in which these things can be expressed. For example, there are languages, known as ‘infinitary’ languages, in which we can form infinite conjunctions ‘ S_1 and S_2 and S_3 and...’. The statement that an even number of things satisfy ‘ Px ’ can be written as an infinite conjunction ‘There is not exactly one thing x such that Px , and there are not exactly three things x such that Px , and there are not exactly five things...’. Infinitary languages can say more than first-order languages; but by the same token it is harder to construct models of theories in infinitary languages, and this makes them less useful for applying model-theoretic arguments in mathematics (see [Infinitary logics](#)). The study of different kinds of formal languages and their expressive power is called ‘generalized model theory’ (see the encyclopedic volume of [Barwise and Feferman 1985](#)).

In 1969 Lindström proved some theorems which compare first-order languages with other languages that have greater expressive power. For each signature S there is a first-order language $L(S)$; it consists of just those formulas which can be built up from the symbols in S using the first-order logical expressions. Suppose that we add some new expressions, for example, quantifiers which express that an even number of things have this or that property. Then each signature S will give rise to a richer language $L'(S)$. One of Lindström’s results is that if the upward and downward Löwenheim-Skolem theorems (quoted in §3 above) hold for ‘elementary equivalence’ in the sense of the languages $L'(S)$ instead of $L(S)$, then for every sentence of $L'(S)$ there is already a sentence of $L(S)$ which has exactly the same models. Roughly speaking, any logic which is at least as strong as first-order logic, but not strong enough to distinguish one infinite cardinal from another, must be first-order logic. Lindström’s theorems show that the choice of first-order languages as the main languages of model theory was not just a happy accident. But they also raise the question of whether there might be purposes for which first-order languages are not the best choice. We may find such purposes outside mathematics.

7 Truth and possible worlds

The interest of model theory for philosophers rests mainly on two items: Tarski's truth definition and Tarski's characterization of logical deduction. Let us begin with the truth definition. (One philosophical text published in 1992 even defines model theory as the programme of giving a truth definition for quantified predicate logic!)

In model-theoretic versions of Tarski's truth definition we define a relation ' $\models$ ' between structures and sentences; ' $M \models \phi$ ' says that the sentence ϕ is true in the structure M . But a typical sentence of English, for example,

(7) Pollution is not the sole cause of the rise in asthma

can be assessed as true or false; no structure is involved.

Some writers contrast absolute truth with model-theoretic truth; the latter is relative to a structure, the former is not. This is a misleading contrast. Truth means the same in model theory as it means anywhere else - the difference lies in the sentences whose truth is being assessed. A typical sentence of a formal language of logic is a schema with schematic variables in it; by itself it does not make a statement. A structure is needed to give meanings to the variables. By contrast, the sentence (7) has no variables, so it needs no structure.

There are some analogues of schematic variables in natural languages. The closest analogues are indexical expressions such as 'I', 'he', 'these' and 'now' (see **Demonstratives and indexicals**). To assess the truth of the sentence

(8) He took those

we need to know who he is and what those are. An answer of the form 'If he is Christopher Naples and those are my football boots, then (8) is true' has almost exactly the same content as the statement that the schema 'X took Y' is true in a certain structure.

Nor does (8) determine the time and place referred to. Maybe the sentence was false yesterday but will be true tomorrow. When we assess its truth, we normally take the default time and place - namely here and now; but any account of the truth or otherwise of (8) needs to make some assumption at least about the time. (Strictly there may also be a question about domains of quantification: in (7) what range of possible causes are we considering?)

Some philosophers go further: we need to know whether (8) is taken to be about this world or some other possible world. Even further: there are sentences such as

(9) It is possible that it will rain tomorrow

whose truth now in this world may depend on what is true in other possible worlds. (Is there a possible continuation of this world in which it will rain tomorrow?)

None of these relativities depends on model theory. But model theory offers a format for handling them. Sentence (8) is true in the structure where the possible world is this one, the time and place are as follows, he is so-and-so and those are such-and-such. Oversimplifying a little, sentence (9) is true in possible world i if and only if there is a world j which is possible relative to i , such that 'It will rain tomorrow' is true in j . Montague (a student of Tarski) offered a truth definition for substantial fragments of English, using devices of exactly these kinds (see **Montague, R.M.**). Similar ideas appear in the semantics of many languages used in philosophy and artificial intelligence; see, for example, **Modal logic**; **Tense and temporal logic**; **Intensional logics**. (See also **Semantics, possible worlds.**)

There may be ontological problems about what a possible world is, but model theory neither adds to nor subtracts from these problems. Most of the structures that one meets in model theory are set-theoretic objects which have nothing to do with possible worlds.

8 Logical truth and consequence

As we saw in the previous section, model theory provides a setting for studying how the truth of a statement is affected if we alter some features of the statement. One traditional way of defining and classifying necessary truths is by noting that they stay true under various kinds of alteration. For example, a necessary truth stays true if we change the possible world; an eternal truth stays true if we change the time; a logical truth stays true if we

systematically reinterpret all the words in it except the logical ones. Classifications along these lines are sometimes described as model-theoretic.

In 1936 Tarski offered just such a definition of logical consequence. On his account, a sentence S is a logical consequence of sentences T if one cannot reinterpret the non-logical words in S and T so as to make S false and all the sentences of T true. He was deliberately vague about which words are logical; he suggested that different choices might lead to different notions of consequence (such as analytic consequence). (See [Consequence, conceptions of §5](#).)

Tarski's definition makes no mention of structures, since he applied it directly to meaningful sentences. To adapt it to a model-theoretic style, we first have to consider formal languages L whose symbols are all either logical expressions or schematic variables. Then, as in §4, if U is a theory and ϕ a sentence in such a language L , we write ' $U \models \phi$ ' to mean that every model of U is a model of ϕ . Finally, we say that S is a logical consequence of T if and only if there are a theory U and a sentence ϕ in some such language L such that $U \models \phi$ and T, S is a substitution instance of U, ϕ . (We say that a sentence S is a 'substitution instance' of a schema ϕ if S comes from ϕ by replacing each schematic variable by a meaningful expression - the same expression at all occurrences of the variable; and likewise for a theory.)

9 Model theory in computer science

Computers are good at handling formulas; they are not so good at set theory. So when model theory appears in computer science, it usually has a supervisory role, for example, to describe what computers or robots are supposed to be doing, or how a computer language is intended to work, or what kind of information a database is capable of retrieving.

There is virtually no place in computer science where the class of all models of a theory has any importance at all; in fact it is hard to see how large uncountable structures could possibly be relevant to computing. But sometimes (a) one is interested in the finite models of some theory. For example, these can represent the possible states of a database, or the possible states of a machine during a computation. Sometimes (b) the relevant models are those in which every element is named by a term; these appear as datatypes, or as representations of the behaviour of a machine through time. Another important case is (c) where one is only interested in a single structure, and the question is which sets and relations can be defined in this structure. The model-theoretic semantics of PROLOG is a case in point: the structure in question is the term algebra, and the relations are defined by PROLOG programs. Computer science is open-ended and fast developing, so that other uses of model theory are very likely to appear.

In the 1980s some theoretical computer scientists started to voice a damaging criticism of first-order model theory. The model-theoretic questions which arise naturally in computer science are to do with defining classes of finite structures, or single structures, or relations within a given structure. The classes are defined in terms of the behaviour of a machine or program. For model theory one needs logical languages in which these same classes can be defined. In practice, first-order languages usually turn out to be grossly unsuitable.

For example, if M is a finite structure representing the information in a database, and the relations of M include a set P and a binary relation R , then it is easy to compute whether the number of elements in P is even, or whether it is possible to get from one given element a to another element b by a chain: $aRc_1R \dots Rc_nRb$. But neither of these questions can be phrased as a first-order sentence. Or, again, the relations defined by a PROLOG program are rarely definable by first-order formulas; this is because the program can perform recursion, so that the relations are built up inductively.

On the other hand, the properties of first-order logic that Lindström pointed to (see §6), such as the Löwenheim-Skolem theorems, are largely irrelevant for computer science.

No definitive replacement for first-order logic has appeared, but computer scientists have begun to develop some very interesting new languages, for example, languages in which one can express the transitive closure of relations. There is also a thriving branch of model theory devoted to finite structures. Here the interesting classes of structure are, for example, the class of structures which give the answer 'yes' to a certain database query, or the class of those structures which can be picked out by an algorithm in a certain complexity class. (See [Complexity, computational; Quantifiers, generalized](#).)

See also: [Logical and mathematical terms, glossary of](#)

WILFRID HODGES

References and further reading

- Barwise, J. and Feferman, S.** (eds) (1985) *Model-Theoretic Logics*, New York: Springer. (A mathematical survey of various logics, unfortunately before there was much model-theoretic work on logics for computer science.)
- Doets, K.** (1996) *Basic Model Theory*, Stanford, CA: CSLI Publications. (A good introduction for readers who know some logic but are not mathematicians.)
- Ebbinghaus, H.-D., Flum, J. and Thomas, W.** (1984) *Mathematical Logic*, Berlin and New York: Springer, 2nd edn, 1994. (A readable general introduction to mathematical logic which emphasizes model theory; includes proofs of Lindström's theorems.)
- Hintikka, J.** (1988) 'On the Development of the Model-Theoretical Tradition in Logical Theory', *Synthese* 77: 1-16. (A philosophical discussion of model theory by a model theorist.)
- Hodges, W.** (1986) 'Truth in a Structure', *Proceedings of the Aristotelian Society* 86: 135-51. (Another philosophical discussion of model theory by a model theorist.)
- Hodges, W.** (1993) *Model Theory*, Cambridge: Cambridge University Press. (A general mathematical text of model theory.)
- Tarski, A.** (1954) 'Contributions to the Theory of Models I', *Indagationes Mathematicae* 16: 572-81. (The paper which first named model theory.)
- Tarski, A.** (1956) *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*, trans. J.H. Woodger, Oxford: Clarendon Press; ed. J. Corcoran, Indianapolis, IN: Hackett Publishing Company, 2nd edn, 1983. (Includes important papers on the conceptual background of model theory, in particular, 'The Concept of Truth in Formalized Languages' (1933), 'On the Concept of Logical Consequence' (1936) and 'On Definable Sets of Real Numbers' (1931).)

Models

Of the many kinds of things that serve as 'models', all function fundamentally as representations of what we wish to understand or to be or to do. Model aeroplanes and other scale models share selected structural properties with their originals, while differing in other properties, such as construction materials and size. Analogue models, which resemble their originals in some aspect of structure or internal relations, are important in the sciences, because they can facilitate inferences about complicated or obscure natural systems. A collection of billiard balls in random motion is an analogue model of an ideal gas; the interactions and motions of the billiard balls are taken to represent - to be analogous to - the interactions and motions of molecules in the gas.

In mathematical logic, a model is a structure - an arrangement of objects - which represents a theory expressed as a set of sentences. The various terms of the sentences of the theory are mapped onto objects and their relations in the structure; a model is a structure that makes all of the sentences in the theory true. This specialized notion of model has been adopted by philosophers of science; on a 'structuralist' or 'semantic' conception, scientific theories are understood as structures which are used to represent real systems in nature. Philosophical debates have arisen regarding the precise extent of the resemblances between scientific models and the natural systems they represent.

1 Types of model

Generally, we use models as representations of various things, including desires, hopes and actions, as well as objects and systems in the real world. Not only do many distinct types of thing serve as models, but they also relate in a variety of ways to the world and to us. Fundamentally, to be a model is to serve in a triadic relation: *a person takes something as a model of something else* (Wartofsky 1979).

One very general meaning of 'model' is as an ideal or paradigm case or exemplar, for example a 'model student', curious and energetic. Here, models serve as ideals towards which people are encouraged to adapt themselves (see [Examples in ethics](#)).

Scale models, for example, model aeroplanes and other miniaturized representations, are objects whose properties and relative proportions resemble their originals. Only some properties, however, are the same in the model and the original; usually the construction materials differ, and the functions of the model and the original differ. Still, scale models exhibit the basic relations which enable one thing to stand in for another; in Charles S. Peirce's terminology, the model is an *icon* which represents the original. Icons are different from symbols in their relationship to the things represented: symbols are tied to their originals purely by convention, and any thing can serve as a symbol for any other thing; icons, in contrast, bear some resemblance or similarity to the things that they represent to us. That is, there must be something about the icon itself which is similar or analogous to the object or system of which it is a model. Thus, scale models are a simple kind of iconic model.

Analogue models are usually more abstract than scale models, and resemble their originals primarily in some aspect of structure or internal relations, rather than in materials or appearance. A collection of billiard balls in random motion is an analogue model of an ideal gas in virtue of certain internal relations and interactions among the billiard balls, and not because any thing in the substance or appearance of gas molecules is *like* billiard balls.

Mathematical models are even more abstract than analogue models, in that the objects and their interrelations in the model are all mathematical entities. In the sciences, mathematical models are used to represent a wide variety of real objects and situations.

Finally, there are metamathematical models, which are technical entities used to understand formal systems, such as mathematical systems themselves.

2 Models in metamathematics

In metamathematics (or 'metalogic', the study of the features of formal systems), a 'model' is a structure that makes all of the sentences in a theory true, where a 'theory' is a set of sentences in a language, the various terms of which are mapped onto objects in the system and their relations. Even in metamathematics, though, 'model' is used in several ways: most generally, a model is a special kind of interpretation, where an interpretation of a

theory consists in both the assignment or mapping of terms in the theory to objects, and to a structure, which consists of objects and their relations to one another. More specifically, an interpretation of a language specifies: (1) a domain (universe of discourse), for example, the range of any variables that occur in any sentence in the language; (2) a designation (denotation, bearer, reference) for each name in the language; (3) a function f , which assigns a value in the domain for any sequence of arguments in the domain, for each function symbol in the language; (4) a truth-value for each sentence letter in the language; and (5) a characteristic function for each predicate letter (Boolos and Jeffrey [1974] 1980). In these cases, models are defined as those interpretations under which all the sentences in the theory are true.

Alternatively, models are simply the structures themselves under a specified mapping assignment (where the mapping assignment is taken to be external to the model; Robinson 1965; Tarski 1941). Models can be any kind of structure; in metalogic, models are usually expressed in terms of set theory, but technically any group of objects and their relations could serve as the model for a theory, provided that it displayed the right structure.

For example, take as a theory the sentences: ‘object A is touching object B’; ‘object C is touching object B’; and ‘object C is not touching object A’ (where A, B and C are terms in the theory, and ‘is touching’ is a relation between two or more of the terms). We can easily construct or imagine a structure which ‘satisfies’ or makes true all of these sentences in the little theory: it could consist of three objects in a row, 1, 2, 3, each one touching only the next. Notice that they could be any kind of object, including cats, jars of jam, and so on, or some mixture of these. The *mapping assignment* for the theory might map A onto 1, B onto 2 and C onto 3. (Equally, it could map A onto 3, B onto 2, and C onto 1.) On the usual logical definition, then, the objects 1, 2, 3 constitute one model of the theory, because a ‘model’ is a structure which can be interpreted so as to make all the sentences in a specific theory true (see **Model theory**).

3 Models in science

Several uses of ‘model’ are standard in the sciences. In each case, models are substitute systems used to investigate and understand the real systems they model. In biology, certain organisms are selected as ‘models’ on the basis of ease of investigation and manipulation; they are investigated intensively, in the hope that the results generalize to other organisms. In biomedical research, for instance, mice are often used as models or stand-ins for studies of drug effects in human beings.

One of the most basic uses of models in physics involves mechanical models of natural processes. Thus, billiard balls in random motion are taken as a model for a gas. This modelling relation does not imply that billiard balls are like gas particles in all respects, simply that gas molecules are analogous to billiard balls. Under the model, some properties of billiard balls ought to be ascribed to gas molecules, that is, motion and impact (the ‘positive analogy’, in Mary Hesse’s terms), while other properties of billiard balls ought not be ascribed to molecules, such as colour and hardness (the ‘negative analogy’). There is also the ‘neutral analogy’, used when we do not know whether the properties are shared; according to Hesse, these properties are what allow us to make new predictions (Hesse 1966: 9). Max Black argued that ‘use of a particular model may...help us to notice what otherwise would be overlooked, to shift the relative emphasis attached to details - in short, to *see new connections*’ (Black 1962: 237). **N.R. Campbell (1920)** claimed that during the development of the kinetic theory of gases, this mechanical billiard ball model of the theory played an essential part in its extension, and thus, that the availability of a working model of a theory is essential to successful theorizing in the sciences. Pierre **Duhem**, in contrast, argued that such use of models in science was preliminary, optional, and potentially misleading, and that proper scientific theories were expressed abstractly and systematically (Duhem 1914).

Many of the philosophical and scientific discussions about the roles of models in science seem to have mixed together several distinct issues. All participants in these debates agree that there must be some analogy between a model used in the sciences (whether mechanical or not) and the phenomena it is being used to explain. One primary concern was how realistically and completely a model - especially a mechanical model - is supposed to represent the aspect of reality under investigation. Hence, Peter Achinstein (1968) suggests a hierarchy of models based on their ontological commitments: at one end, we have models which are simply supposed to provide possible mechanisms for how natural systems might be operating, while at the other end, we have concrete claims that the real world is thoroughly like the entities and dynamics in the model. In the latter case, the choice of model amounts to a metaphysical commitment regarding the contents of the universe - which things and relations really

exist. A great deal of discussion has thus centred around which of these attitudes - from purely instrumental to strongly ontological - scientists and philosophers should take to the models they use.

The degree to which a model is taken as exactly and adequately representing reality is one of the most significant ways that claims about models in science can differ. What is frequently called an ‘instrumental’ use of models amounts to treating the models as calculating devices; slightly more committed, ontologically, are idealized models, which can be seen as either false-but-handly simplifications of processes in the natural system being discussed, or as approximately true representations of some of the forces operating in the natural system (see [Idealizations](#)).

One link between the metamathematical and scientific uses of models lies in the notion of interpretation. In the sciences, models are sometimes used as tools for making sense of theories which are not otherwise immediately comprehensible; a model can thus help the community of scientists articulate and pursue scientific theories. The wave and particle interpretations of light and of quantum mechanics are well-known examples; both the wave and particle models offer ways to understand the theoretical (and well-confirmed) equations, but they present different and incompatible features (see [Quantum mechanics, interpretations of](#)).

Finally, models are generally acknowledged to be psychologically valuable to scientists; they serve the heuristic functions of helping scientists envisage very complex systems, and they help simplify inferences about those systems. Even models that are known to be misrepresentations of the real world may sometimes possess these virtues (Wimsatt [1987](#)).

4 Models in philosophy of science

While discussion of the above uses of models in the sciences occupied philosophers of science for many decades, one of their basic assumptions was that models are essentially different from scientific theories. This distinction has come under scrutiny since the 1950s, and much work in recent philosophy of science going under the name of ‘structuralist’ or ‘semantic’ approaches has centred on analysing scientific theories in terms of metamathematical models.

For the middle third of the twentieth century, the reigning logical positivist approaches to science understood scientific theories as ‘sets of deductively connected sentences in a formal language’ combined with rules for interpreting some of the terms of that language. In contrast to this linguistic view of theories, some philosophers advocated viewing theories as structures, ‘which are propounded as standing in some representational relationship to actual and physically possible phenomena’ (Suppe [1979: 320](#)). More specifically, scientific theories are understood as presenting models in the metamathematical sense, that is, arrangements of objects and their relations. These structures function as iconic models for the scientists using them: they characterize, in idealized circumstances, the systems that they represent.

Take evolutionary population genetics as an example. Population geneticists tend to present their theories in the form of mathematical models. This means that, given the mathematical models, it is possible to examine the structures that instantiate that theory. Under this ‘semantic approach’ to theories, the focus is on the structures themselves, rather than on an attempt to reconstruct, in some theoretical language, the sentences of the theory, as demanded by the axiomatic and positivist approaches to theory structure. (It remains a matter of debate whether there is an important difference between viewing a theory as a set of sentences and viewing it as a set of models. Some philosophers have suggested that the difference is merely pragmatic or heuristic: scientists sometimes find it more natural to regard a theory one way rather than another, but the two views are in principle intertranslatable. See Schaffner [1993: 99-125](#).)

This use of metamathematical models is an extension from their use in interpreting formal systems. When models serve as icons, some aspects of the model’s entities and their interrelations are taken to be similar to the phenomena in the natural world that are being investigated. In metamathematics, the more precise notion of isomorphism is central to evaluating the similarity of models. Two models are isomorphic if it is possible to make a one-to-one mapping from each element in one model to an element in the other model, and to make a one-to-one mapping from each relation among elements in one model to a relation among corresponding elements in the other model. While the notion of similarity between model and nature may be intuitive, precision demands that the natural system be represented, abstracted or measured in some fashion, if evaluations of isomorphism between the

theoretical model and nature itself are to be attempted. Typically, measurements or data from the natural system are arranged into a 'data model', which provides the basic entities and relations to which the theoretical models are compared (Suppes 1962).

Many questions can be raised about the relations between scientific theories - when understood as models - and the real world. For instance, it is possible that there are relations or entities that appear in the models that have no correspondence in the real world. Alternatively, there may be mechanisms represented in the model which are hypothesized to represent the way the natural world really works; the model might then be interpreted as an approximation or idealization of the real world systems, including mechanisms that we cannot observe directly. Debates among philosophers about 'scientific realism' and 'antirealism' have centred on the issues of what sorts of inference are necessary, or are justified, by the successful application of a scientific model to the real world (Churchland and Hooker 1985; van Fraassen 1980; Giere 1988; Suppe 1989).

See also: [Scientific method](#); [Scientific realism and antirealism](#); [Theories, scientific](#)

ELISABETH A. LLOYD

References and further reading

- Achinstein, P.** (1968) *Concepts of Science*, Baltimore, MD: Johns Hopkins University Press. (Referred to in §3. Careful analysis of various conceptions of models and their uses in the sciences.)
- Black, M.** (1962) *Models and Metaphors: Studies in Language and Philosophy*, Ithaca, NY: Cornell University Press. (Referred to in §3. Insightful discussion of philosophical understanding of models and the uses of models in scientific thought; not very technical.)
- Boolos, G. and Jeffrey, R.C.** (1972) *Computability and Logic*, New York: Cambridge University Press; 2nd edn 1980. (Second edition referred to in §2. Standard text for metalogic. Requires familiarity with predicate logic.)
- Burks, A.W.** (1949) 'Icon, Index, and Symbol', *Philosophical and Phenomenological Research* 9: 673-89. (Review of Peirce's semiotics, good explication of the concept of 'icon' proposed by Peirce.)
- Campbell, N.R.** (1920) *Physics, the Elements*, Cambridge, Cambridge University Press; repr. as *Foundations of Science*, New York: Dover, 1957. (Referred to in §3. Controversial early advocacy of the centrality of models in scientific research.)
- Carnap, R.** (1942) *Introduction to Semantics*, Cambridge, MA: Harvard University Press. (Provides more background and discussion than many other basic texts. Useful introductory chapters and examples of metamathematical models.)
- Churchland, P.M. and Hooker, C.W.** (1985) *The Image of Science*, Chicago, IL: University of Chicago Press. (Referred to in §4. Useful collection of essays addressing the interpretation of models in science.)
- Duhem, P.** (1914) *The Aim and Structure of Physical Theory*; trans. P.P. Wiener, Princeton, NJ: Princeton University Press, 1954, 2nd edn. (Referred to in §3. Influential criticism of the role of mechanical models in physics.)
- Fraassen, B.C. van** (1980) *The Scientific Image*, Oxford: Clarendon Press. (Referred to in §4. Influential discussion of models in science. Defence of empiricist interpretation of scientific theories.)
- Giere, R.** (1988) *Explaining Science: A Cognitive Approach*, Chicago, IL: University of Chicago Press. (Referred to in §4. Accessible discussion and applications of the semantic conception of scientific theories. Excellent introductory chapters.)
- Hesse, M.B.** (1966) *Models and Analogies in Science*, Notre Dame, IN: University of Notre Dame Press. (Referred to in §3. Original and influential discussion of the scientific and logical uses of models.)
- Peirce, C.S.** (1934) *Collected Papers of Charles Sanders Peirce*, vol. II, Cambridge, MA: Harvard University Press. (Original, foundational work on semiotics, icons, symbols; fairly challenging.)
- Robinson, A.** (1965) *Introduction to Model Theory and to the Metamathematics of Algebra*, Amsterdam: North Holland. (Referred to in §2. Authoritative text, primarily technical.)
- Schaffner, K.F.** (1993) *Discovery and Explanation in Biology and Medicine*, Chicago, IL: University of Chicago Press. (Referred to in §4. Discusses the semantic view and alludes to the debate about differences between the semantic and syntactic views.)
- Suppe, F.** (1977) *The Structure of Scientific Theories*, Urbana, IL: University of Illinois Press, 2nd edn. (Excellent review of scientific and philosophical discussions concerning models. Comparison of logical positivist and semantic conceptions of theory structure. Outstanding bibliography.)
- Suppe, F.** (1979) 'Theory Structure', *Current Research in Philosophy of Science*, East Lansing, MI: Philosophy of Science Association. (Quoted in §4; useful brief review of some central issues in philosophy of science.)

- Suppe, F.** (1989) *The Semantic Conception of Theories and Scientific Realism*, Urbana, IL: University of Illinois Press. (Referred to in §4. Best source for history and context of treatments of models in philosophy of science. Careful discussion of issues regarding how extensively scientific theories represent the natural world.)
- Suppes, P.** (1962) 'Models of Data', in E. Nagel, P. Suppes and A. Tarski (eds) *Logic, Methodology, and Philosophy of Science: Proceedings of the 1960 International Congress*, Stanford, CA: Stanford University Press. (Referred to in §4. Sophisticated but accessible analysis of the extraction and organization of quantitative evidence from nature, and of its role in theory testing.)
- Suppes, P.** (1967) 'What is a Scientific Theory?', in S. Morgenbesser (ed.) *Philosophy of Science Today*, New York: Basic Books. (Influential defence of the use of metamathematical models in philosophy of science. Clear and accessible.)
- Tarski, A.** (1941) *Introduction to Logic and to the Methodology of the Deductive Sciences*, New York: Oxford University Press; 2nd edn, 1946. (Referred to in §2. Presentations of formal logic and models by one of the developers of metamathematics.)
- Wartofsky, M.** (1979) *Models: Representation and the Scientific Understanding*, Dordrecht: Reidel. (Referred to in §1. Original essays discussing the metaphysical, epistemological and social significance of various types of model. Excellent and readable.)
- Wimsatt, W.** (1987) 'False Models as Means to Truer Theories', in N. Nitecki and A. Hoffman (eds) *Neutral Models in Biology*, London: Oxford University Press. (Referred to in §3. Insightful discussion of a variety of attitudes towards models in the sciences.)

Multiple-conclusion logic

Ordinary arguments can have any number of premises but only one conclusion. Multiple-conclusion logic also allows for any number of conclusions in an argument, regarding them as setting out the field of possibilities among which the truth must lie if the premises are true. Such an argument counts as valid if it is impossible for all the premises to be true and all the conclusions false. Anything that can be said about premises can now be said, *mutatis mutandis*, about conclusions, and much of the interest of the subject comes from exploiting this duality. Putting conclusions on a par with premises reflects the idea that truth and falsity, and likewise acceptance and rejection, are polar notions standing on a par with one another.

1 Multiple-conclusion arguments

Multiple conclusions can best be understood by analogy with multiple premises. Premises function collectively: a number of premises taken together have a strength which far exceeds the sum of their separate strengths. The very existence of deductive logic turns on this fact, which is not undermined by the fact that a finite number of premises are always equivalent to a single one, namely their conjunction. For this equivalence itself relies on the working of the rule 'From A and B infer $A \& B$ ', understood as involving two separate premises (not one conjunctive one), so it does nothing to establish the redundancy of arguments with more than one premise.

Multiple conclusions also function collectively: a set of conclusions will typically be deducible although no single conclusion is deducible. They are best thought of disjunctively, as setting out the field within which the truth must lie on the assumption that the premises are all true. Just as enlarging a set of premises strengthens them (makes more things follow from them), so enlarging a set of conclusions weakens them (makes them follow from more things): as the field containing the truth becomes larger it becomes less informative. Of course, a finite number of conclusions are equivalent to a single one, namely their disjunction, but this no more makes multiple conclusions redundant than the analogous fact about conjunction makes multiple premises redundant.

The definition of logical consequence or implication is now that a set of premises implies a set of conclusions if it is impossible for all the premises to be true and all the conclusions false. The transitivity of implication is a key property of it, which in conventional single-conclusion logic takes the form of the 'cut' principle; that if a set of premises Γ implies a conclusion A , and if Γ, A imply B , then Γ implies B . The corresponding multiple-conclusion principle takes this novel form: if Γ, Θ_1 implies Θ_2 , Δ for every partition of the set Θ into two disjoint parts Θ_1 and Θ_2 , then Γ implies Δ .

Logicians often depict arguments by diagrams in which a horizontal line indicates that the sentence immediately below is inferred from those immediately above, so that the diagram branches upwards from one conclusion. Multiple-conclusion arguments can be represented in a similar fashion, except that now the graph can branch downwards as well as upwards. So (1) shows one variety of 'proof by cases', where one starts with a premise of the form ' A or B ' and presents the alternatives as distinct cases as a preliminary to establishing the conclusion by separate, parallel arguments for case A and case B .

$$(1) \quad \begin{array}{c} \frac{A \vee B}{A \quad B} \\ \vdots \quad \vdots \\ C \quad C \end{array} \quad (2) \quad \begin{array}{c} \frac{A \vee B}{A \quad B} \\ \hline A \& B \end{array}$$

(1) is untypically simple, for in general the combination of upward and downward branching means that branches which have split may subsequently rejoin, creating circuits. Moreover, it turns out that it is essential to allow for circuits (or some equivalent sophistication) if one is to obtain an adequate variety of arguments. But some circuits are vicious, (2), for example. Much of the mathematical development of the subject has therefore been concerned with finding criteria for a diagram to have a valid pattern, in the sense that any argument which has the same bare layout (that is, abstracting altogether from the sentences involved and paying attention only to the way the steps are arranged relative to one another) is bound to be valid provided its component steps are individually valid. This is a good example of a new concept which, because patterns of single-conclusion argument are so rudimentary by comparison, is scarcely appreciable within the conventional framework.

2 Motivation and applications

Multiple-conclusion logic commends itself to the attention of the philosophical logician because putting conclusions on a par with premises is a natural corollary of acknowledging that truth and falsity are polar notions which stand on a par with one another, and likewise acceptance and rejection. Deduction is no longer tilted towards acceptance alone: an argument regarded as going from premises Γ to conclusions Δ with an eye to acceptance can equally be regarded as going from Δ to Γ with an eye to rejection. The connection can be stated in suitably symmetrical terms, namely that if the argument is valid it is irrational to accept all of Γ and reject all of Δ .

For the mathematical logician the removal of the asymmetry embodied in single-conclusion logic means that some anomalies disappear, and distinctions and results which needed qualification become unconditional. Recall, for example, the conventional way of developing a system of logic. First the semantics is set out, then implication is defined in semantic terms and finally, where possible, it is axiomatized, that is, shown to be equivalent to formal deducibility according to some system of rules of inference. This exposes a deficiency. Evidently the notion of implication can be used to say things about the assignments of truth-values determined by the semantics - for example, that none of them makes certain premises true and a certain conclusion false. But the information that can be conveyed in this way is only a fraction of the whole. For example, there is no way of saying in terms of implication alone whether the set of all sentences is satisfiable or not. And even if all the notions with which logicians work - logical truth, implication, satisfiability and so on - are taken together, the situation is similar. What we would surely like, but do not have, is a logical notion or notions capable of saying all there is to say about the semantics at the level appropriate to sentences.

In a similar vein, Carnap (1943) pointed out that, although one may manage to axiomatize a logic initially given semantically, one can never work back and retrieve the semantics from the axiomatization. For example, any axiomatization which is sound and complete for classical propositional logic will equally be sound and complete with respect to a nonstandard semantics in which, in addition to the classically possible assignments of truth-values, every sentence (and its negation) is simultaneously satisfiable. In short, there is no way, in purely formal deductive terms, of expressing the fact that ‘ $\sim$ ’ signifies negation. The situation is akin to the non-categoricity of theories, whereby a first-order theory of real numbers will also have a nonstandard interpretation in a merely countable domain (see [Löwenheim-Skolem theorems and nonstandard models](#)). Those who are rightly concerned about the implications of that notorious result might do well to reflect that things are no better for negation.

Multiple-conclusion logic remedies both deficiencies. Implication is now an adequate vehicle for saying all there is to say about any semantics at the level of sentences, for every fact of the form ‘It is (is not) possible for these sentences Γ to be true and those ones Δ false’ is expressible by the statement that Γ does not (does) imply Δ . And by using multiple-conclusion rules one can obtain axiomatizations from which the intended semantics can be uniquely reconstructed. For example, one can work back from the rules ‘From A , $\sim A$ infer’ and ‘Infer A , $\sim A$ ’ to the classical truth table for ‘ $\sim$ ’.

Multiple-conclusion rules of inference can also provide an elegant alternative to the often complicated rules involving assumptions and their discharge, which characterize natural deduction systems. (1) above illustrates the idea by using a multiple-conclusion rule in place of the natural-deduction rule of $\vee$ -elimination. A profound application of this idea can be found in Gentzen’s work on ‘sequents’ if one interprets his sequents metalogically as statements about multiple-conclusion implication, though this was not how Gentzen himself regarded them (see [Natural deduction, tableau and sequent systems](#)).

See also: [Logical and mathematical terms, glossary of](#)

TIMOTHY SMILEY

References and further reading

- Carnap, R.** (1943) *Formalization of Logic*, Cambridge, MA: Harvard University Press. (The first explicit appearance of the subject.)
- Kneale, W.C.** (1956) ‘The Province of Logic’, in H.D. Lewis (ed.) *Contemporary British Philosophy*, 3rd series, London: Allen & Unwin; abridged in W. Kneale and M. Kneale, *The Development of Logic*, Oxford: Clarendon

Press, 1962, ch. 9, §3. (Another pioneering piece, with a very readable introduction.)

Shoesmith, D.J. and Smiley, T.J. (1978) *Multiple-Conclusion Logic*, Cambridge: Cambridge University Press. (A systematic treatise, including a discussion of the history of the subject.)

Natural deduction, tableau and sequent systems

Different presentations of the principles of logic reflect different approaches to the subject itself. The three kinds of system discussed here treat as fundamental not logical truth, but consequence, the relation holding between the premises and conclusion of a valid argument. They are, however, inspired by different conceptions of this relation. Natural deduction rules are intended to formalize the way in which mathematicians actually reason in their proofs. Tableau systems reflect the semantic conception of consequence; their rules may be interpreted as the systematic search for a counterexample to an argument. Finally, sequent calculi were developed for the sake of their metamathematical properties.

All three systems employ rules rather than axioms. Each logical constant is governed by a pair of rules which do not involve the other constants and are, in some sense, inverse. Take the implication operator ' $\rightarrow$ ', for example. In natural deduction, there is an introduction rule for ' $\rightarrow$ ' which gives a sufficient condition for inferring an implication, and an elimination rule which gives the strongest conclusion that can be inferred from a premise having the form of an implication. Tableau systems contain a rule which gives a sufficient condition for an implication to be true, and another which gives a sufficient condition for it to be false. A sequent is an array $\Gamma \vdash \Delta$, where Γ and Δ are lists (or sets) of formulas. Sequent calculi have rules for introducing implication on the left of the ' $\vdash$ ' symbol and on the right.

The construction of derivations or tableaux in these systems is often more concise and intuitive than in an axiomatic one, and versions of all three have found their way into introductory logic texts. Furthermore, every natural deduction or sequent derivation can be made more direct by transforming it into a 'normal form'. In the case of the sequent calculus, this result is known as the cut-elimination theorem. It has been applied extensively in metamathematics, most famously to obtain consistency proofs. The semantic inspiration for the rules of tableau construction suggests a very perspicuous proof of classical completeness, one which can also be adapted to the sequent calculus. The introduction and elimination rules of natural deduction are intuitionistically valid and have suggested an alternative semantics based on a conception of meaning as use. The idea is that the meaning of each logical constant is exhausted by its inferential behaviour and can therefore be characterized by its introduction and elimination rules.

Although the discussion below focuses on intuitionistic and classical first-order logic, various other logics have also been formulated as sequent, natural deduction and even tableau systems: modal logics, for example, relevance logic, infinitary and higher-order logics. There is a gain in understanding the role of the logical constants which comes from formulating introduction and elimination (or left and right) rules for them. Some authors have even suggested that one must be able to do so for an operator to count as logical.

1 Natural deduction

Systems of natural deduction were first devised in the late 1920s by Gerhard Gentzen (1935), a student of Hilbert's, and, independently, by Stanislaw Jaśkowski. Their derivations formalize proofs from assumptions and are most conveniently represented as tree-like structures with the conclusion at the root and an assumption at the top of each branch. Assumptions may be discharged by the application of a rule in the course of the derivation, and the conclusion will depend only on the undischarged assumptions. Here is one formulation of the introduction and elimination rules for the usual logical operators:

- $$\begin{array}{ll}
 (1) \quad \frac{A \quad B}{A \wedge B} & (2) \text{ (a) } \frac{A \wedge B}{A} \\
 & \text{ (b) } \frac{A \wedge B}{B} \\
 (3) \text{ (a) } \frac{A}{A \vee B} & (4) \frac{[A] \quad [B] \quad A \vee B \quad C}{C} \\
 \text{ (b) } \frac{B}{A \vee B} & \\
 (5) \quad \frac{[A] \quad B}{A \rightarrow B} & (6) \quad \frac{A \quad A \rightarrow B}{B} \\
 (7) \quad \frac{\perp}{\neg A} & (8) \quad \frac{A \quad \neg A}{\perp} \\
 (9) \quad \frac{A(a)}{\forall x A(x)} & (10) \quad \frac{\forall x A(x)}{A(a)} \\
 & [A(a)] \\
 (11) \quad \frac{A(a)}{\exists x A(x)} & (12) \quad \frac{\exists x A(x) \quad C}{C}
 \end{array}$$

Rule (3a), for example, adds a new root to the tree, transforming a derivation of A from unspecified assumptions into a derivation of $A \vee B$ from those same assumptions. Introductions are written on the left, eliminations on the right. ‘ $\perp$ ’ is a constant for falsity. A formula enclosed in square brackets indicates that assumptions of this form in the derivation of the premise displayed below it may be discharged by the inference in question.

In rules (9) and (12), a cannot occur in $A(x)$ and is called the ‘proper parameter’ of the inference; it must satisfy the following restrictions. In (9), a cannot occur in any assumption on which the premise depends; in (12), a cannot appear in C nor in any assumption upon which the premise C depends except for those of the form $A(a)$. Because they allow the discharge of assumptions, it is better to view applications of the rules $\rightarrow$ -introduction and $\vee$ - and $\exists$ -elimination as operating on whole derivations rather than simply on their conclusions. A similar remark can be made about $\forall$ -introduction because it requires that the derivation of $A(a)$ in its premise satisfy an extra condition. This system of introductions and eliminations constitutes minimal logic. A system equivalent to Heyting’s axiomatization of intuitionistic logic is obtained by adding a rule which allows any formula to be inferred from a falsehood.

$$\frac{\perp}{A}$$

Classical logic can be obtained by adding to the intuitionistic system as axioms all instances $A \vee \neg A$ of the law of the excluded middle, or a rule such as the following.

$$\frac{\neg \neg A}{A}$$

The major premise of an elimination is the formula whose principal operator is displayed in the statement of the rule. A ‘maximum formula occurrence’ in a derivation is both the conclusion of an introduction and the major premise of an elimination. It constitutes a needless detour and can be removed. This can be illustrated in the case of implication by the following pair of derivations (where Π_1 and Π_2 are derivations of A and B , respectively):

$$\begin{array}{c}
 [A] \\
 \Pi_2 \\
 \frac{\Pi_1}{A} \quad \frac{B}{A \rightarrow B} \\
 \hline
 B
 \end{array}$$

The one on the right is obtained from Π_2 by replacing each occurrence of the assumption A discharged by the

application of $\rightarrow$ -introduction by a copy of the derivation Π_1 of A . If no assumption is discharged, the derivation on the left is just replaced by Π_2 . In either case, the result is a more direct derivation of the same conclusion from the same or fewer assumptions. A transformation of this kind is called a ‘reduction’ step, and similar reductions can be defined for the other operators. The special form of the elimination rules for ‘ $\vee$ ’ and ‘ $\exists$ ’ give rise to what are called ‘maximum segments’. These consist of consecutive occurrences of the same formula such that the first is the conclusion of an introduction, the last is the major premise of an elimination and all except the first are conclusions of an application of $\vee$ - or $\exists$ -elimination. To remove such segments it is necessary to permute applications of $\vee$ - and $\exists$ -elimination with other inferences.

A derivation is said to be ‘normal’ if it contains no maximum formula occurrence or segment. Gentzen’s normal form theorem states that for every derivation in the intuitionistic system, there is a normal one of the same conclusion from the same or fewer assumptions. It is proved by prescribing a method for systematically removing all detours from a derivation, and thus a stronger result is established - sometimes called the ‘normalization theorem’ - namely, that every derivation Π can be *reduced* to one in normal form. The reduction steps can be so chosen that every sufficiently long sequence of them beginning with Π terminates in a unique normal derivation. This is called the strong normalization theorem. Roughly speaking, a normal derivation is constructed by first breaking up its assumptions using eliminations and then building up its conclusion using introductions. That is why every formula appearing in the derivation will be a subformula of an assumption or the conclusion. All three forms of the theorem have been exploited to yield various proof-theoretic results for formalized mathematical theories. In addition, the significance of the normalization procedure itself has been studied in a more general context. It has been argued that the steps required to reduce a derivation to normal form do not change its meaning. In consequence, derivations which reduce to the same normal form have the same meaning or express the same proof. More controversially, the converse has also been claimed to hold.

2 The sequent calculus

The theorems discussed at the end of the previous section do not extend straightforwardly from the intuitionistic to the classical version of natural deduction. By exploiting the definability of disjunction and existential quantification in classical logic, however, they can be shown to hold for a system of natural deduction adequate for classical logic. A rather more sophisticated argument is required to establish them for a classical natural deduction calculus in which all the usual logical constants are taken as primitive. Gentzen, however, did not extend his theorem to the classical system described above. He chose rather to reformulate it as a calculus of sequents. It is a trivial matter to rewrite the introductions and eliminations as sequent rules by placing the assumptions on which a formula depends on the left-hand side of the sequent. The rules for ‘ $\rightarrow$ ’, for example, will look as follows.

$$\frac{A, \Gamma \vdash B}{\Gamma \vdash A \rightarrow B} \qquad \frac{\Gamma \vdash A \rightarrow B \quad \Delta \vdash A}{\Gamma, \Delta \vdash B}$$

Although assumptions have been eliminated in favour of axioms of the form $A \vdash A$, nothing else of significance has been changed. In fact, the resulting calculus is just a notational variant of natural deduction. Gentzen’s idea was to rewrite only the introduction rules in this way, and to replace the eliminations by new rules for introducing a formula on the left side of a sequent. These left rules are displayed below. (a cannot occur anywhere in the conclusion of the last rule.)

$$\begin{array}{c}
 \frac{A, \Gamma \vdash C}{A \wedge B, \Gamma \vdash C} \quad \frac{B, \Gamma \vdash C}{A \wedge B, \Gamma \vdash C} \\
 \frac{A, \Gamma \vdash C \quad B, \Gamma \vdash C}{A \vee B, \Gamma \vdash C} \\
 \frac{\Gamma \vdash A \quad B, \Delta \vdash C}{A \rightarrow B, \Gamma, \Delta \vdash C} \\
 \frac{\Gamma \vdash A}{\neg A, \Gamma \vdash} \\
 \frac{A(a), \Gamma \vdash C}{\forall x A(x), \Gamma \vdash C} \\
 \frac{A(a), \Gamma \vdash C}{\exists x A(x), \Gamma \vdash C}
 \end{array}$$

The rules may be read in different ways. Their connection with natural deduction is best brought out by interpreting the symbol ‘ $\vdash$ ’ as derivability. Then the last of them, for example, asserts that a derivation of C from $A(a), \Gamma$ can be transformed into one of C from $\exists x A(x), \Gamma$ (subject to the restriction on a).

In addition, Gentzen introduced some structural rules: ‘interchange’ for permuting formulas on a list, ‘contraction’ for amalgamating two occurrences of the same formula, and ‘thinning’ or ‘weakening’ for adding formulas to a list. Because an empty list on the right of a sequent can be interpreted as falsity, there is no longer any need for the constant ‘ $\perp$ ’. Furthermore, the intuitionistic negation rule can now be derived because thinning on the right justifies the inference from $\Gamma \vdash$ to $\Gamma \vdash A$. In the presence of the structural rules, it does not matter whether the left rule for ‘ $\vee$ ’ is formulated as above or as follows.

$$\frac{A, \Gamma \vdash C \quad B, \Delta \vdash C}{A \vee B, \Gamma, \Delta \vdash C}$$

And likewise for the other two premise rules, both left and right. Recently J.-Y. Girard, in his work on linear logic (see [Girard 1995](#)), has embarked upon a searching investigation of the proof-theoretic significance of the structural rules, and of the consequences of omitting or weakening them in various ways. In their absence, he distinguishes two versions of disjunction, for example: an additive and a multiplicative one according to which formulation of its left rule is adopted.

3 Comparing sequents and natural deduction

It is easy to translate a derivation of the sequent $\Gamma \vdash A$ into a derivation in natural deduction of A from assumptions included in the list Γ - into a normal one, in fact. Call this derivation its $\mathcal{N}$ translation. Applications of right rules translate directly into introductions. The situation with left rules is only slightly more complicated. Consider conjunction, for example, and suppose that Π is the $\mathcal{N}$ translation of a derivation of $A, \Gamma \vdash C$, then the required derivation of C from assumptions in $A \wedge B, \Gamma$ is obtained by substituting copies of the two-line derivation

$$\frac{A \wedge B}{A(\top)}$$

for the appropriate occurrences of the assumption A in Π .

Converting natural deduction derivations into sequent ones is less straightforward. Introductions correspond as before to applications of a right rule. To translate eliminations, however, we need not only the corresponding left rule, but also a rule called ‘cut’.

$$\frac{\Gamma \vdash A \quad A, \Delta \vdash C}{\Gamma, \Delta \vdash C}$$

With the addition of this rule every sequent derivation will still have an $\mathcal{N}$ translation, although we can no longer be sure that it will be normal. The $\mathcal{N}$ translation of a derivation which ends with the cut displayed above is obtained by substituting copies of the $\mathcal{N}$ translation of the derivation of its left premise for the appropriate occurrences of the assumption A in the $\mathcal{N}$ translation of the derivation of its right one.

To see why cut is needed for translating the elimination rules, consider again the case of conjunction and suppose that we have already succeeded in translating a derivation Π of $A \wedge B$ from assumptions Γ into a derivation δ of the sequent $\Gamma \vdash A \wedge B$. Then what needs to be shown is that δ can be converted into a derivation of $\Gamma \vdash A$. This is accomplished by a cut applied to the conclusions of δ and a derivation whose $\mathcal{N}$ translation is $(\dagger)$ above, as shown in the following figure.

$$\frac{\delta \quad \frac{A \vdash A}{A \wedge B \vdash A}}{\Gamma \vdash A \wedge B \quad A \wedge B \vdash A} \Gamma \vdash A$$

This translation exploits the fact that the result of applying $\wedge$ -elimination to Π is the same as substituting Π for the assumption $A \wedge B$ in the derivation $(\dagger)$ above. The other elimination rules are handled similarly.

Clearly, every formula appearing in the premise(s) of a left or right rule is a subformula of a formula appearing in its conclusion, and the same holds for the structural rules. Consequently, any derivation constructed only by the use of these rules will be such that it contains only subformulas of formulas occurring in its final sequent.

Gentzen's normal form theorem in this context takes the form of a cut-elimination theorem, and he was able to prove that every derivation can be transformed into a cut-free derivation of the same sequent. His proof, although similar in kind to the one for natural deduction, does involve additional complications. Furthermore, cut-free forms are not in general unique. It would appear therefore that little advantage has been gained by passing from natural deduction to a sequent calculus, especially since sequent derivations are for the most part less intuitive.

In favour of sequent derivations is the fact that the absence of assumptions makes them easier to work with from a metamathematical standpoint. To verify the correctness of a sequent derivation it is enough to check it locally, that is, to look at each sequent and its immediate successors individually. There are no global conditions on the derivation that need to be satisfied (as, for example, that a certain parameter cannot appear in any assumption). This also makes coding derivations by means of Gödel numbers much easier. A more significant advantage, however, concerns the relationship between the calculus described above and classical logic. Gentzen observed that, by the simple expedient of allowing a list containing more than one formula to appear on the right-hand side of a sequent, he could obtain a system adequate for classical logic. The classical versions of the right and left rules for ' $\rightarrow$ ', for example, read as follows.

$$\frac{A, \Gamma \vdash \Delta, B}{\Gamma \vdash \Delta, A \rightarrow B} \quad \frac{\Gamma \vdash \Delta, A \quad B, \Psi \vdash \Phi}{A \rightarrow B, \Gamma, \Psi \vdash \Delta, \Phi}$$

All the remaining rules, including cut, are similarly generalized; no new rules need to be added. To see that the resulting calculus is adequate for classical logic, note that the sequent $\vdash A \vee \neg A$ is now easily derivable. Applying $\neg$ -right to $A \vdash A$ yields $\vdash A, \neg A$, from which the desired sequent follows by two applications of $\vee$ -right and a couple of structural inferences. It is possible therefore to give a single proof of the cut-elimination theorem for both classical and intuitionistic systems. The proof for the classical system involves no new ideas, and it specializes immediately to a proof for the intuitionistic one. One need simply observe that each reduction prescribed in the proof yields an intuitionistic derivation when applied to one.

Gentzen's theorem has been fruitfully applied to yield a variety of results in proof theory. The Craig interpolation theorem, for example, follows from it easily, as do various decidability results. Gentzen himself utilized the method of cut elimination to prove the consistency of formal number theory, that is, of a system obtained by augmenting the sequent calculus with appropriate arithmetical principles. (In fact, his original proof was for a theory based on natural deduction, but he subsequently rewrote it for the sequent calculus.) His methods have since been adapted and extended, in many different ways and to a variety of different situations. A version of the cut-elimination theorem for second-order logic, for example, has yielded a consistency proof for analysis, formalized as a theory of second-order arithmetic (see [Proof theory §4](#)). The information provided by results such as these depends upon the methods required to establish them, and their significance is understood differently in different foundational traditions. Nevertheless, there is no doubt that the methods pioneered by Gentzen have been one of the mainstays of proof theory in the tradition of Hilbert, as it has developed in the wake of Gödel's theorems.

Although cut-free derivations are in some sense simpler and more direct than their counterparts containing cuts, the elimination procedure described by Gentzen increases the length of a derivation each time a cut is removed

from it. Furthermore, this turns out to be an essential feature of any cut-elimination procedure. The same holds true for the elimination of maximum formula occurrences from natural deduction derivations and, as one might expect, the efficiency of the tableau method described below is correspondingly increased when an analogue to the cut rule is added to the system. This simple fact has given rise to a great deal of research on the complexity of derivations, and very refined bounds on the length of cut-free ones have been established for a variety of systems and theories. The subject of complexity has attracted particular attention from researchers interested in implementing automated reasoning systems (see [Complexity, computational](#)).

4 Tableau systems

The discussion of the previous sections was based on an interpretation of the sequent calculus as characterizing the relation of derivability under assumptions in natural deduction. With the passage from intuitionistic to classical sequents, this interpretation is no longer possible. The classical rules are easily seen to be correct under a semantic interpretation, however, according to which $\Gamma \vdash \Delta$ is taken to mean that for every model of Γ there is at least one formula in Δ that is true in it. If the rules are read backwards, from conclusion to premise(s), they represent steps in the search for a counterexample to $\Gamma \vdash \Delta$, a model in which all the formulas in Γ are true, while all those in Δ are false. To find a counterexample to $\Gamma \vdash \Delta, A \wedge B$, for example, we may search in one of two directions, either for a counterexample to $\Gamma \vdash \Delta, A$ or for one to $\Gamma \vdash \Delta, B$. The first person to exploit this feature of the rules seems to have been E.W. Beth. It underlies the tableau system for classical logic which he formulated in the mid-fifties (see his later work of 1964 for details). However, the general idea of a no-counterexample treatment of logic is also to be found in the works of other logicians of the period, including J. Hintikka, S. Kanger and K. Schütte.

A semantic tableau for the sequent $\Gamma \vdash \Delta$ begins as a figure consisting of two columns, with the formulas in Γ listed at the head of the left one and the formulas in Δ at the head of the right. The rules extend the tableau by operating on individual formulas in these lists. A left rule operates on A in Γ and allows us to enter into the tableau subformulas of A in such a way that the truth of the left entries together with the falsity of the right ones together constitute a sufficient condition for A to be true. Similarly, a right rule operates on A in Δ and allows us to enter subformulas whose truth and falsity are sufficient for A to be false. As before, there are pairs of rules for each of the logical constants, and some of these rules require the tableau to split into a pair of subtableaus, each with its own left and right column. Since all the classical connectives and quantifiers are definable in terms of ' $\neg$ ', ' $\rightarrow$ ' and ' $\exists$ ', it will be sufficient to list tableau rules for just these three operators:

If $\neg A$ appears in the left column, write A in the right one.

If $\neg A$ appears in the right column, write A in the left one.

If $A \rightarrow B$ appears in the left column, split the tableau into two subtableaus, write B in the left column of one subtableau and $\neg A$ in the right column of the other.

If $A \rightarrow B$ appears in the right column, write B below it and $\neg A$ in the left column.

If $\exists x A(x)$ appears in the left column, write $A(a)$ in the left column, where a is a new parameter.

If $\exists x A(x)$ appears in the right column, write $A(a)$ in the right column, where a is any parameter that has already been introduced into the tableau (or a new one if none has been introduced).

The left rule for ' $\rightarrow$ ', for example, reflects the fact that in classical logic there are two ways to verify an implication: its antecedent may be false or its consequent true. The right rule reflects the fact that there is only one way to refute one: its antecedent must be true and its consequent false. In all cases except the last, the formula to which the rule has been applied is cancelled to indicate that the rule cannot be applied to it again. In the case of right ' $\exists$ ', the possibility of applying the rule with respect to a different parameter is left open. A subtableau is closed if there is at least one formula which occurs in both its left and right column. A tableau is closed if all its subtableaus are closed. $\Gamma \vdash \Delta$ is derivable if it has a closed tableau. Finally, a tableau has been terminated if it is not closed, but no further rules are applicable to any of its subtableaus.

By way of example, the following closed tableau establishes the validity of the sequent

$$\neg \exists x \neg (A(x) \rightarrow B) \vdash \exists x A(x) \rightarrow B.$$

$\neg\exists x\neg(A(x) \rightarrow B)$	$\exists x A(x) \rightarrow B$
$\exists x A(x)$	B
$A(a)$	
	$\exists x\neg(A(x) \rightarrow B)$
	$\neg(A(a) \rightarrow B)$
$A(a) \rightarrow B$	
¹	² B
¹ $A(a)$	²

Each line is obtained from a previous one by applying one of the rules for tableau construction. Notice that the initial tableau splits into two subtableaus after line six. (The first and third cells of line seven are the left and right columns of one, the second and fourth of the other.) Both subtableaus are closed, the one because $A(a)$ appears on the left in line three and on the right in line seven, and the other because B appears on the right in line two and on the left in line seven.

The rules for tableau construction are really nothing other than upside-down versions of the classical sequent rules. Indeed, one need do little more than invert a closed tableau to convert it into a cut-free derivation in the sequent calculus. There is no need to augment the tableau rules by a rule analogous to cut because their deductive adequacy can be demonstrated directly without reference to the derivations of any other system. The rules may be applied to a sequent in any order (although it is convenient to postpone applying right ' $\exists$ ' for as long as possible) and three possibilities can arise:

- (1) The tableau closes after a finite number of steps, in which case the disjunction of the formulas in Δ is a classical consequence of the formulas in Γ .
- (2) The tableau is terminated after a finite number of steps. This will be because every uncanceled formula (other than existential ones to which right ' $\exists$ ' is no longer applicable) is atomic. It is then easy to construct a model in which each atomic formula appearing on the left of an open subtableau is true and each one appearing on the right is false. Such a model turns out to be a counterexample to the original sequent in the sense that all the formulas in Γ are true in it while those in Δ are false.
- (3) The tableau is never terminated. This possibility arises because right ' $\exists$ ' may be applied repeatedly to a formula as needed. Nevertheless, as Beth showed, we can prove generally that a counterexample can also be constructed from any such infinite tableau.

The upshot is that the rules of tableau construction provide a complete proof procedure for classical logic. Despite their semantic motivation, it should be emphasized that these are purely syntactic rules - and hence that there is a cut-free derivation of every derivable sequent in the sequent calculus as well. This semantic approach to the cut-elimination theorem is less elementary than Gentzen's original proof. More significantly, however, the connection between a derivation and its cut-free form is lost.

Because tableaus may split into subtableaus, the method is cumbersome to use. A considerable simplification was effected by Raymond Smullyan, who introduced the idea of 'signed' formulas. Given $\Gamma \vdash \Delta$, the formulas in Γ receive the sign T and those in Δ the sign F. A tableau now begins with a single list of signed formulas, and left and right rules are reformulated as rules for formulas with the signs T and F respectively. The result is a single tree, rather than a series of tableaus, which branches instead of splitting. A closed branch will contain an occurrence of some formula A with the sign T and another occurrence with the sign F, and a tree is closed when every branch is closed. (Alternatively, one can dispense with signs and form a single list from the formulas in Γ together with the negations of those in Δ .) It is in this form, rather than the original one, that the tableau method has found its way into the literature. An analogous modification of the sequent calculus was introduced by Kurt Schütte. Here the effect is to make derivations less intuitive, although the calculus is more concise and better suited for some applications.

Just as natural deduction seems to be a more appropriate system for intuitionistic logic, so tableaus seem better adapted to express classical consequence. Nevertheless, as Beth showed, they can be modified to yield a proof procedure for intuitionistic logic. Subsequently, tableau techniques were extended to modal systems, by Kripke among others, and what are now called Kripke models have their origin in constructions deriving from these.

See also: [Logical and mathematical terms, glossary of](#)

A.M. UNGAR

References and further reading

- Beth, E.W.** (1964) *The Foundations of Mathematics*, Amsterdam: North Holland; New York: Harper & Row, 2nd edn, 1966. (§§67-70, 144-5 include accounts of tableaux for classical and intuitionistic logic, respectively; one of the few places where tableaux are presented in their original two-column form.)
- Boolos, G.** (1984) 'Don't Eliminate Cut', *Journal of Philosophical Logic* 13: 373-8. (Includes a striking illustration of just how much shorter than its cut-free counterpart a derivation with cut can be.)
- Došen, K.** (1993) 'A Historical Introduction to Substructural Logics', in K. Došen and P. Schroeder-Heister (eds) *Substructural Logics*, Oxford: Clarendon Press, 1-30. (A brief and readable survey of the motivations for and effects of restricting structural rules in sequent calculi, with a useful bibliography.)
- Fitting, M.** (1983) *Proof Methods for Modal and Intuitionistic Logics*, Dordrecht: Reidel. (A thorough and self-contained presentation of natural deduction, sequent and tableau methods for a variety of logics, including classical.)
- Gentzen, G.** (1935) 'Untersuchungen über das logische Schließen', *Mathematische Zeitschrift* 39: 176-210, 405-565; repr. Darmstadt: Wissenschaftliche Buchgesellschaft, 1969; trans. in *The Collected Papers of Gerhard Gentzen*, ed. M.E. Szabo, Amsterdam and London: North Holland, 1969, 68-131. (Gentzen's original, and very readable, account of sequent and natural deduction systems for classical and intuitionistic logic.)
- Girard, J.-Y.** (1995) 'Linear Logic: Its Syntax and Semantics', in J.-Y. Girard, Y. Lafont and L. Regnier (eds) *Advances in Linear Logic*, Cambridge: Cambridge University Press, 1-42. (An introductory, albeit challenging, account of linear logic by its originator.)
- Kleene, S.C.** (1967) *Mathematical Logic*, New York: Wiley. (Chapter 4 describes a very direct proof of the completeness of the classical cut-free sequent rules.)
- Prawitz, D.** (1965) *Natural Deduction: A Proof-Theoretical Study*, Uppsala: Almqvist & Wiksell. (A thorough but elementary presentation of natural deduction and the normalization theorem.)
- Prawitz, D.** (1971) 'Ideas and Results in Proof Theory', in J.E. Fenstad (ed.) *Proceedings of the Second Scandinavian Logic Symposium*, Amsterdam: North Holland, 235-307. (A survey of some proof-theoretic applications of natural deduction.)
- Schwichtenberg, H.** (1977) 'Proof Theory: Some Applications of Cut-Elimination', in J. Barwise (ed.) *Handbook of Mathematical Logic*, Amsterdam: North Holland, 867-95. (A survey of some proof-theoretic applications of the sequent calculus; requires more background than the previous item.)
- Stålmarck, G.** (1991) 'Normalization Theorems for Full First Order Classical Natural Deduction', *Journal of Symbolic Logic* 56 (1): 129-49. (This article is rather technical.)
- Sundholm, G.** (1983) 'Systems of Deduction', in D. Gabbay and F. Guenther (eds) *Handbook of Philosophical Logic*, Dordrecht: Reidel, vol. 1, 133-88. (A more detailed survey of the systems discussed in this article, with a useful bibliography; recommended as the next thing to read.)
- Sundholm, G.** (1986) 'Proof Theory and Meaning', in D. Gabbay and F. Guenther (eds) *Handbook of Philosophical Logic*, Dordrecht: Reidel, vol. 3, 471-506. (Another well-written survey article; explains how natural deduction and related systems may contribute to a theory of meaning which does not depend on the notion of truth.)

Necessary truth and convention

Necessary truths have always seemed problematic, particularly to empiricists and other naturalistically-minded philosophers. Our knowledge here is a priori - grounded in appeals to what we can imagine or conceive (or can prove on that basis) - which seems hard to reconcile with such truths being factual, short of appealing to some peculiar faculty of a priori intuition. And what mysterious extra feature do necessary truths possess which makes their falsity impossible? Conventionalism about necessity claims that necessary truths obtain by virtue of rules of language, such as that 'vixen' means the same as 'female fox'. Because such rules govern our descriptions of all cases - including counterfactual or imagined ones - they generate necessary truths ('All vixens are foxes'), and our a priori knowledge is just knowledge of word meaning. Opponents of conventionalism argue that conventions cannot ground necessary truths, particularly in logic, and have also challenged the notion of analyticity (truth by virtue of meaning). More recent claims that some necessary truths are a posteriori have also fuelled opposition to conventionalism.

1 The linguistic theory of necessity

Many of the clearest necessary truths - such as 'Squares are rectangles' - seem simply to reflect the meaning of the words involved: we need not postulate some metaphysical essence of squareness which the necessity of 'Squares are rectangles' reflects. Thus, the position common to conventionalists is that necessary truths are explained by conventional rules of language, rather than a mind-independent modal structure of the world. For this reason, the general conventionalist position has often been called the linguistic theory of necessity (or, of the a priori). This was the dominant view about necessity from the 1930s-50s, particularly under the influence of logical positivism (see **Logical positivism**). Drawing largely upon claims they found in Wittgenstein's *Tractatus* (see below), the positivists rejected Kant's claim that there are synthetic a priori truths, trumpeting the slogan 'The a priori is analytic' (see **Analyticity**). What exactly analyticity was - beyond, in some way, truth by virtue of the meanings of words - was not entirely clear, which gave rise to different versions of the linguistic theory. But for most, conventionalism itself was beyond doubt and, apart from the few explicitly engaged in trying to spell out just how conventions explained necessity, most writers used different ways of articulating the view ('All necessary truths are tautologies', 'Necessity simply reflects linguistic rules') interchangeably simply to invoke the linguistic theory.

An early expression of conventionalism is found in Locke's proposal that all essences are merely nominal - they simply reflect the meaning of the subject term, as in 'Gold is a metal'. In this, Locke meant to contradict the then-received view that essences are determined by a mind-independent sorting of things. Similarly, he says essences 'are made by the mind' (1689: III.vi.26), and since 'kinds' - or in Locke's terms, 'sorts' - are the kinds they are because of their essences, Locke says the sorts themselves are 'the Workmanship of the Understanding' (1689: III.vi.12) (see **Locke, J. §4**). This well exemplifies the metaphysically deflationary thrust of conventionalism.

While there are various conventionalist views, they agree negatively both in their suspicion of appeals to imagination and conceivability as sources of factual knowledge, and in their finding the notion of necessary facts - and a distinction between essential and accidental properties - rather mysterious (see **Essentialism**). On the positive side, they agree in seeing appeals to imagination and conceivability as appeals to what we would say in various imagined situations, which they see as a way to find out what rules of language we are implicitly employing.

The version of conventionalism presented in **Wittgenstein's *Tractatus*** focuses on the notion of 'tautology', that is, truth-functional compound statements, such as 'If it is raining, then it is raining', which are true for all assignments of truth-values to the propositions from which they are built (1922: 4.46). Wittgenstein argues that tautologies say nothing about the world; rather, the symbols, given their meanings, themselves determine the truth of such statements (1922: 4.461-4661, 5.142). He claims that all truths of logic are tautologies (1922: 6.1-112) and seems to hold this for all a priori truths (for example, 1922: 5.134, 5.634). There are necessary truths which are not truths of logic, such as 'All bachelors are unmarried'. However, since 'bachelor' means 'unmarried man', the statement, when analysed, says, 'All unmarried men are unmarried', the form of which is 'Whatever is *F* and *G* is *F*', which is a tautology. One problem this version of conventionalism faced was accounting for necessary truths such as 'Nothing can be simultaneously both red and green all over' - for it is hard to see how to assign meanings to 'red'

and 'green' whereby this can be reduced to a tautology.

Other conventionalists have concentrated more on the general idea that truths such as 'All squares are rectangles' express rules of language. Their 'truth by convention' could perhaps serve as a model for all necessary truths. One version of this idea was that 'Necessarily, all squares are rectangles' simply means 'It is a rule of language that "square" applies to something only if "rectangle" also applies'; **A.J. Ayer** wrote that necessary propositions 'simply record our determination to use words in a certain fashion' (1936: 84). This view was regularly criticized for identifying necessary a priori truths with contingent empirical statements about a language or its speakers. A subtler variant was that one should not really speak of necessary truths at all - what look to be assertions are really, covertly, prescriptions ('Do not apply "square" to anything you would not apply "rectangle" to') or perhaps decisions of some sort. One problem here is that necessary statements can stand in logical relations, can figure essentially in truth-functional compounds, and so on, which seems to require that they have truth-values, unlike prescriptions. The most common version of this view is simply that the rule determines the truth of the sentence; given the rule, nothing can count as a square but not a rectangle. This view does not require that all rules have the form of definitions which allow us, via substitution, to reduce all necessary truths to syntactical tautologies; a rule prohibiting the simultaneous use of 'red all over' and 'green all over' to an object need not be rooted in a full, necessary-and-sufficient-conditions style definition of some constituent term (see **Carnap 1952**).

While many attacks on conventionalism only properly address particular versions, certain problems challenge any sweeping conventionalism. By the late 1950s, not only was conventionalism no longer dogma but, to many, seemed simply mistaken. One central problem concerns the fact that whether some convention obtains is always a contingent matter - but how can something which is contingent make anything necessary? For instance, even conventionalists (generally) acknowledge that different conventions would not affect the necessity of: if if p then q , and p , then q . But if different conventions would not change what is necessary, how can the conventions be responsible? Different conventions can change which proposition is expressed - but not the necessity of the proposition actually expressed. Another general problem concerns the idea that logical truths obtain by convention. The conventionalist wants to say that, given certain rules, it follows that certain statements are true (and necessary). But 'follows' is an expression of necessitation. **Quine (1966)** and others argue that in order to derive the truths of logic from conventions, one must apply logical rules - one must presuppose logic in order to derive logic. Additionally, attacks on the analytic/synthetic distinction, particularly vivid in arguments put forward by **Putnam (1975)** that apparently analytic truths are actually revisable empirical claims ('Cats are animals'), cast further doubt on the analytic explanation of necessity. If there are no analytic truths, they cannot explain necessity (see **Analyticity**).

2 Necessity a posteriori

In *Naming and Necessity* (1980), **Saul Kripke** argued that the supposition that all necessary truths are a priori is mistaken. His examples included empirical identity statements, such as 'Hesperus is Phosphorus', statements of material origin, such as 'This lectern was originally made of (a particular piece of) wood' and statements of kind essence and identity, such as 'Gold has atomic number 79'. Since these are not a priori, they cannot simply be true by convention, and if conventionalism was already on the wane, Kripke's examples, and similar ones from **Putnam (1975)**, seemed to kill it decisively. Because these truths are not logically necessary, but are still necessary absolutely, they are sometimes called 'metaphysically necessary'.

However, it has been noted that even if these necessary truths are not true by convention, they may still owe their necessity to convention (see **Coppock 1984**, **Sidelle 1989**). The arguments for proposed necessary a posteriori truths each involve commitment to some general principle, for instance, that if a material object originates in bit of matter M , then it essentially originates in M . These principles - in so far as the arguments for the necessities are plausible - are plausibly a priori, for the arguments rest on the familiar appeals to what we can imagine, or would be willing to say. The conventionalist proposal, then, is that these general principles are analytic, and that while these conventions do not determine which non-modal statements are true (what bit of matter something did originate in), they are responsible for these truths being necessary (the principle determines that nothing with a different origin can count as this object). So conventionalism does not require that all necessary truths are *true* by convention, but only that this is why they are necessary.

See also: [Conventionalism](#); [Essentialism](#) [Language, conventionality of](#); [Logical and mathematical terms, glossary of](#)

ALAN SIDELLE

References and further reading

- Ayer, A.J.** (1936) *Language, Truth and Logic*, London: Gollancz; repr., New York: Dover, 2nd edn, 1946, 16-18 and ch. 4. (Influential and often cited presentation of conventionalism about necessity (and a priori truth); seen by many as claiming that necessary truths say that there is a convention (but see Broad 1936).)
- Broad, C.D.** (1936) 'Are There Synthetic A Priori Truths?', *Proceedings of the Aristotelian Society*, supplement: 102-16. (Expresses scepticism about Ayer's conventionalism.)
- Carnap, R.** (1947) *Meaning and Necessity*, Chicago, IL: University of Chicago Press, esp. chaps 1 and 5. (A rigorous attempt to formulate a system in which it is clear how necessary truth arises from the language.)
- Carnap, R.** (1952) 'Meaning Postulates', *Philosophical Studies* 3: 65-73. (Introduces a device for trying to account for non-tautological necessary truths. Also appears in Carnap 1947.)
- Coppock, P.** (1984) 'Review of N. Salmon, *Reference and Essence*', *Journal of Philosophy* 81: 261-70. (Sketches a conventionalist account of necessary a posteriori truths.)
- Ewing, A.C.** (1940) 'The Linguistic Theory of A Priori Propositions', *Proceedings of the Aristotelian Society*: 207-44. (Classic criticism of conventionalism.)
- Dummett, M.** (1959) 'Wittgenstein's Philosophy of Mathematics', *Philosophical Review* 68: 324-48. (Critical discussion of Wittgenstein's conventionalism (1956).)
- Forbes, G.** (1985) *The Metaphysics of Modality*, Oxford: Clarendon Press, ch. 9. (Defends modal concepts against scepticism and psychological accounts, and offers a broadly conventionalist view.)
- Fraassen, B.C. van** (1977) 'The Only Necessity is Verbal Necessity', *Journal of Philosophy* 74: 71-85. (Argues against any irreducibly natural necessity.)
- Kneale, W.** (1946) 'Truths of Logic', *Proceedings of the Aristotelian Society*: 207-34. (Argues against a conventionalist view of logic.)
- Kneale, W.** (1947) 'Are Necessary Truths True By Convention?', *Proceedings of the Aristotelian Society*, supplement: 118-33. (Forceful negative argument, emphasizing the conflict between 'necessary' and 'conventional', and critically discussing the conventionalist's appeal to alternative logics and algebras.)
- Kripke, S.A.** (1980) *Naming and Necessity*, Cambridge, MA: Harvard University Press. (Argues for the conceptual separation of necessity and a prioricity, and for necessary a posteriori truths.)
- Lewis, D.K.** (1969) *Convention*, Cambridge, MA: Harvard University Press. (Argues that sense can be made of the idea of non-explicit conventions, contra Quine 1953.)
- Locke, J.** (1689) *An Essay concerning Human Understanding*, ed. P.H. Nidditch, Oxford: Clarendon Press, 1979. (Referred to in §1. Locke's discussion of substance in III.vii classically expounds his conventionalism about essence and kinds.)
- Malcolm, N.** (1940) 'Are Necessary Propositions Really Verbal?', *Mind* 49: 189-203. (Representative expression of conventionalism, particularly concerned to avoid the claim that necessary statements express the proposition that some convention holds, or state any fact.)
- Pap, A.** (1958) *Semantics and Necessary Truth*, New Haven, CT: Yale University Press, esp. ch. 7. (An extended argument against conventionalist views of necessity.)
- Putnam, H.** (1975) *Philosophical Papers II: Mind, Language and Reality*, Cambridge: Cambridge University Press, 1979. ('Is Semantics Possible?' (referred to in §1) argues against the traditional view of meaning upon which analyticity rests; development of [Quine 1953](#). 'The Meaning of "Meaning"' (referred to in §2) argues for necessary a posteriori status for certain claims about natural kinds, such as 'Tigers are mammals' and 'Water is H₂O'.)
- Quine, W.V.** (1953) 'Two Dogmas of Empiricism', in *From a Logical Point of View: Nine Logico-Philosophical Essays*, Cambridge, MA: Harvard University Press, 2nd edn, 1980. (Most often cited argument against the analytic-synthetic distinction, and so against truth by convention.)
- Quine, W.V.** (1966) 'Truth By Convention', in *The Ways of Paradox and Other Essays*, Cambridge, MA: Harvard University Press. (Referred to in §1. A classic argument against accounting for truth by convention alone.)
- Schlick, F.A.M.** (1931) 'Gibt es ein materiales Apriori?', trans. 'Is There a Factual A Priori?', in H. Feigl and W.S. Sellars (eds) *Readings in Philosophical Analysis*, New York: Ridgeview, 1982. (Highly representative example of assertion that necessity and a prioricity can only derive from linguistic convention.)

- Sidelle, A.** (1989) *Necessity, Essence and Individuation: A Defense of Conventionalism*, Ithaca, NY: Cornell University Press.(Referred to in §2. Defends a conventionalist account of necessary a posteriori truths.)
- Wittgenstein, L.J.J.** (1922) *Tractatus Logico-Philosophicus*, trans. C.K. Ogden and F.P. Ramsey, London: Routledge; trans. D.F. Pears and B.F. McGuinness, London: Routledge, 1961.(Highly influential source of the view that all necessary truths are tautologies and that tautologies say nothing about the world.)
- Wittgenstein, L.J.J.** (1956) *Remarks on the Foundations of Mathematics*, Oxford: Blackwell.(Critical remarks about the necessity involved in mathematical truths and proofs. Helpful interpretation in Wright (1980) and Dummett (1959).)
- Wright, C.** (1980) *Wittgenstein on the Foundations of Mathematics*, London: Duckworth, and Cambridge, MA: Harvard University Press, esp. Part 3.(Sympathetic interpretation of Wittgenstein's conventionalism about necessity.)
- Yablo, S.** (1993) 'Is Conceivability a Guide to Possibility?', *Philosophy and Phenomenological Research* 53: 1-42.(Tries to defend the objectivity of our a priori methods of investigating modality, thus responding to the epistemological arguments for conventionalism.)

Non-constructive rules of inference

For some theoretical purposes, generalized deductive systems (or, 'semi-formal' systems) are considered, having rules with an infinite number of premises. The best-known of these rules is the ' ω -rule', or rule of infinite induction. This rule allows the inference of $\forall n\Phi(n)$ from the infinitely many premises $\Phi(0), \Phi(1), \dots$ that result from replacing the numerical variable n in $\Phi(n)$ with the numeral for each natural number. About 1930, in part as a response to Gödel's demonstration that no formal deductive system had as theorems all and only the true formulas of arithmetic, several writers (most notably, Carnap) suggested considering the semi-formal systems obtained, from some formulation of arithmetic, by adding this rule. Since no finite notation can provide terms for all sets of natural numbers, no comparable rule can be formulated for higher-order arithmetic. In effect, the ω -rule is valid just in case the relevant quantifier can be interpreted substitutionally; looked at from the other side, the validity of some analogue of the ω -rule is the essential mathematical characteristic of substitutional quantification.

1 The ω -rule

In a standard formalized deductive system, the axioms form a decidable set of formulas (or, under Gödel-numbering, a recursive set of numbers), and each rule is a decidable relation between (conclusion) formulas and finite sets of (premise) formulas. (A set, or relation, is 'decidable' if there is an algorithm for determining whether a given item belongs to it, or given items stand in it.) A proof in the system can therefore be thought of as a tree of formulas, with the theorem to be proved at the root, axioms at the leaves, and at each node a formula which follows by a rule from its predecessors.

For some theoretical purposes, however, 'semi-formal' systems (generalized deductive systems) are considered, having rules with an infinite number of premises. The best-known of these rules is that called the rule of infinite induction, or the ' ω -rule'. This rule allows the inference of $\forall n\Phi(n)$ from the infinitely many premises $\Phi(0), \Phi(1), \dots$ that result from replacing the numerical variable n in $\Phi(n)$ with the numeral for each natural number. About 1930, in part as a response to Gödel's demonstration that no formal deductive system had as theorems all and only the true formulas of arithmetic, several writers (most notably, Carnap) suggested considering the semi-formal systems obtained, from some formulation of arithmetic, by adding this rule.

An axiomatic system enriched with the ω -rule can, obviously, prove all the truths of first-order arithmetic: without loss of generality we may think of the atomic sentences of arithmetic as equations between terms made up out of numerals and the signs for addition and multiplication, so propositional logic and elementary algorithms (which can be simulated in standard axiomatic systems) suffice to prove all quantifier-free sentences of arithmetic, after which any true existential quantification can be inferred by the standard rule of existential quantifier introduction from one of its true instances, and any true universal quantification by the ω -rule from the set of all its true instances. (Proofs of true 'prenex' sentences (that is, those made up of a string of quantifiers followed by an atomic sentence) have this form exactly; other truths can be deduced from their prenex equivalents by standard means.)

If the language of the system goes beyond first-order arithmetic the situation is more complicated. Within a few years of its proposal, Rosser (1937) showed that the addition of the ω -rule to a standard system of higher-order arithmetic did not suffice to render the system complete. There are, however, some partial completeness results. Schütte (1960) showed that standard axioms and rules enriched with the ω -rule remain complete when the language of first-order arithmetic is supplemented with free variables for sets of numbers. Earlier, Orey (1956) had shown that any set of sentences of second-order arithmetic consistent with the ω -rule has an ' ω -model', that is, a model in which the second-order quantification (quantification over sets of numbers) can have a nonstandard interpretation in the manner of Henkin (1950), but in which the first-order quantifiers range over the genuine natural numbers. (Note that Schütte's result is that every sentence true in the standard model is a theorem of a certain theory, whereas Orey's is more analogous to completeness theorems for logics in specifying only a kind of model.)

The finite trees of formulas counting as proofs in a given formal system themselves form a decidable set of arrays, and can be Gödel-numbered; the set of (Gödel-numbers of) theorems of the system is thus defined by a Σ_1^0 formula, that is, a formula obtained from one expressing a recursive relation by existentially quantifying over

numbers (see [Recursion-theoretic hierarchies](#)). This is the general form of a definition of a recursively enumerable set, and, as shown most elegantly by Smullyan (1961), any recursively enumerable set can be represented by the set of theorems of a formal system. A proof in a semi-formal system can be represented by a set of numbers, so it is tempting to define the set of (Gödel numbers of) theorems of a semi-formal system by a formula starting with an existential quantifier over sets, by analogy with the definition of a recursively enumerable set by existential quantification over numbers.

But an ω -proof is a *well-founded* tree of formulas: it may have infinitely many branches, though each branch is only finitely long. So specifying just which sets of numbers correspond to infinite proofs would require a second quantifier over sets, and this would yield a Σ_2^1 formula as definiens. It is better to define the set of theorems directly, without reference to the infinite proofs, as the set of formulas belonging to every set which contains the axioms, is closed under the ordinary rules of inference, and contains a universal (numerical) quantification if it contains all its instances. This yields a Π_1^1 formula (a formula obtained by prefixing existential quantifiers over sets to a formula containing at most first-order quantifiers) as a definiens. Since Frege (1879) it has been standard to use a formula of this form in giving an explicit definition of an inductively defined notion, and sets definable by Π_1^1 formulas are sometimes called ‘inductive’ sets (see Aczel (1977) or Moschovakis (1974) for details). Any Π_1^1 set can be represented by the set of theorems of a semi-formal system.

Formal systems of arithmetic can be given an elegant formulation analogous to natural deduction or sequent-calculus formulations of pure logic, with the principle of mathematical induction taking the form of a rule of inference similar to the rules for the quantifiers. When this is done, it turns out that Gentzen’s cut-elimination theorem for pure logic (1935) does not extend in its general form to arithmetic, but his later proof of the consistency of arithmetic can be seen as proceeding via a special case of cut-elimination that does hold. Kurt Schütte noted that it was possible to modify Gentzen’s methods to prove a general cut-elimination theorem for various systems of arithmetic, provided that the cut-free proofs are allowed to employ the ω -rule. He and his colleagues have developed this into a useful tool for analysing formal systems of arithmetic: the complexity of the cut-free ω -proofs corresponding to proofs in a particular formal system turns out to be a measure of the strength of that system, correlating in particular with the class of number-theoretic functions whose totality can be proved in the system.

2 Related rules

Stronger infinitary rules, producing semi-formal systems of second-order arithmetic which are still incomplete but are stronger than those formed by adding the ω -rule to standard formal systems, have been studied by Mostowski (1961) and Enderton (1967). No satisfactory analogue of the ω -rule for introducing the universal quantifications over sets in classical higher-order arithmetic or set theory is possible because a formal language cannot contain terms for all the uncountably many sets in the standard model, but higher-order analogues of the ω -rule have been studied in the context of non-classical set theories. Schütte (1960) describes a type-free theory inspired by Ackermann; and Fitch, in publications starting in the late 1940s, obtained results closely related to those of later workers on the inductive semantics of type-free theories by studying semi-formal systems with rules similar to the ω -rule.

In the other direction, various weaker forms of the ω -rule have been studied. It is sometimes possible to prove in a finitary metatheory (or in a theory like Peano arithmetic used as its own syntactic metatheory) that all the instances of a certain quantification are provable in a formal system. A rule allowing the inference of a universal (numerical) quantification from the statement that all its instances are provable is sometimes called the ‘formalized’ ω -rule. Such a rule, unlike those discussed above, is of finitary character; added to a formalized system of arithmetic it amounts to a kind of reflection principle. Feferman (1962) showed that every true sentence of first-order arithmetic is provable in some formal system in a transfinite hierarchy of systems obtained from a standard axiomatic system by iterated addition of reflection principles. (Much earlier, Hilbert (1931) had proposed a restricted version of the formalized ω -rule in what seems to have been an attempt to get around Gödel’s incompleteness theorems.)

A somewhat related idea is the ‘constructive’ ω -rule: take $\forall n\Phi(n)$ to be a theorem if there is a recursive function assigning to each natural number (the Gödel number of) a formal proof of the corresponding instance of $\Phi(n)$. Despite the name, this is a non-constructive rule, since it is not in general a decidable matter whether or not there is a recursive function of the sort required: Shoenfield (1959) shows that all truths of first-order arithmetic are

provable with it.

A.P. HAZEN

References and further reading

- Aczel, P.** (1977) 'An Introduction to Inductive Definitions', in J. Barwise (ed.) *Handbook of Mathematical Logic*, Amsterdam: North Holland, 739-82. (A survey of results on inductive definability; shorter than Moschovakis (1974).)
- Carnap, R.** (1935) 'Ein Gültigkeitskriterium für die Sätze der klassischen Mathematik', *Monatshefte für Mathematik und Physik* 42: 163-90; incorporated in *The Logical Syntax of Language*, trans. A. Smeaton, London: Kegan Paul, 1937. (Uses the ω -rule to give a definition of mathematical 'validity'.)
- Dunn, J.M. and Belnap, N.D.** (1968) 'The Substitution Interpretation of the Quantifiers', *Noûs* 2: 177- 85. (An influential study of the notion of substitutional quantification, pointing out that analogues of the ω -rule are valid for it.)
- Enderton, H.** (1967) 'An Infinitistic Rule of Proof', *Journal of Symbolic Logic*, 32: 447-51. (Treats a non-constructive rule more nearly analogous to the ω -rule than, but weaker than, that studied in Mostowski (1961).)
- Feferman, S.** (1962) 'Transfinite Recursive Progressions of Axiomatic Theories', *Journal of Symbolic Logic* 27: 259-316. (Study of iterated addition of reflection principles.)
- Feferman, S.** (1984) 'Toward Useful Type-Free Systems I', *Journal of Symbolic Logic* 49: 75-111. (Includes a historical survey describing the Ackermann-Schütte and Fitch type-free systems.)
- Frege, G.** (1879) *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*, Halle: Nebert; trans. 'Begriffsschrift, a Formula Language, Modelled Upon That of Arithmetic, for Pure Thought', in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967, 1-82. (First formalization of second-order logic; replacement of recursive definitions with explicit definitions quantifying over sets.)
- Gentzen, G.** (1935) 'Untersuchungen über das logische Schließen', *Mathematische Zeitschrift* 39: 176- 210, 405-565; repr. Darmstadt: Wissenschaftliche Buchgesellschaft, 1969; trans. 'Investigations into Logical Deduction', *American Philosophical Quarterly* 1: 288-306, 2: 204-18, 1964; repr. in *The Collected Papers of Gerhard Gentzen*, ed. M.E. Szabo, Amsterdam and London: North Holland, 1969, 68-131. (Cut elimination.)
- Gentzen, G.** (1936) 'Die Widerspruchsfreiheit der reinen Zahlentheorie', *Mathematische Annalen* 112: 493- 565; trans. 'The Consistency of Elementary Number Theory', in *The Collected Papers of Gerhard Gentzen*, ed. M.E. Szabo, Amsterdam and London: North Holland, 1969, 132-213. (Proof of consistency of arithmetic.)
- Henkin, L.** (1950) 'Completeness in the Theory of Types', *Journal of Symbolic Logic* 15: 81-91. (Introduces nonstandard models for higher-order logic and arithmetic.)
- Hilbert, D.** (1931) 'Die Grundlegung der elementaren Zahlenlehre', *Mathematische Annalen* 104: 485-94. (An early proposal of a restricted version of the formalized ω -rule.)
- Moschovakis, Y.N.** (1974) *Elementary Induction on Abstract Structures*, Amsterdam: North Holland. (Inductive definition.)
- Mostowski, A.** (1961) 'Formal Systems of Analysis Based on an Infinitistic Rule of Proof', in *Infinitistic Methods: Proceedings of Symposium on Foundations of Mathematics*, Warsaw: PWN, and New York, Oxford, London, Paris: Pergamon Press, 141-66; repr. in *Foundational Studies* vol. 1, Amsterdam: North Holland, 1979. (Investigates systems of second-order arithmetic validated by a certain subclass of ω -models.)
- Orey, S.** (1956) 'On ω -Consistency and Related Properties', *Journal of Symbolic Logic* 21: 246-56. (Completeness theorem.)
- Pohlers, W.** (1987) 'Contributions of the Schütte School', appendix in G. Takeuti, *Proof Theory*, Amsterdam: North Holland, 2nd edn. (Survey of work by Schütte and colleagues.)
- Pohlers, W.** (1989) *Proof Theory: An Introduction*, Berlin: Springer. (Introductory textbook on the work of the Schütte school.)
- Rosser, J.** (1937) 'Gödel Theorems for Non-Constructive Logics', *Journal of Symbolic Logic* 2: 129- 37. (Incompleteness theorem.)
- Schmerl, U.** (1982) 'Iterated Reflection Principles and the ω -Rule', *Journal of Symbolic Logic* 47: 721-33. (Relates ideas of Feferman (1962) to more explicitly stated rules.)
- Schütte, K.** (1960) *Beweistheorie*, Berlin: Springer; trans. J.N. Crossley, *Proof Theory*, Berlin and New York: Springer, 1997; also trans. with modifications, S. Toledo, *Tableau Systems for First Order Number Theory and Certain Higher Order Theories*, Berlin: Springer, 1975 (Treatise on proof theory. Toledo's book is a 'pseudo-translation' of the first edition, helpfully substituting Beth- Smullyan tableaux for Schütte's

idiosyncratic formulation of basic logic.)

Shoenfield, J. (1959) 'On a Restricted ω -Rule', *Bulletin de l'Académie Polonaise des Sciences* 7: 405- 7.(Result on effective ω -rule.)

Smullyan, R. (1961) *Theory of Formal Systems*, Princeton, NJ: Princeton University Press.(Elegant presentation of basic recursion theory in terms of simple formal systems.)

Tarski, A. (1933) 'Einige Betrachtungen über die Begriffe ω -Widerspruchsfreiheit und der ω -Vollständigkeit', *Monatshefte für Mathematik und Physik* 40: 97-112; trans. 'Some Observations on the Concepts of ω -Consistency and ω -Completeness', in *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*, ed. J.H. Woodger, Oxford: Oxford University Press, 1956.(Introduces the ω -rule.)

Non-monotonic logic

A relation of inference is ‘monotonic’ if the addition of premises does not undermine previously reached conclusions; otherwise the relation is non-monotonic. Deductive inference, at least according to the canons of classical logic, is monotonic: if a conclusion is reached on the basis of a certain set of premises, then that conclusion still holds if more premises are added.

By contrast, everyday reasoning is mostly non-monotonic because it involves risk: we jump to conclusions from deductively insufficient premises. We know when it is worthwhile or even necessary (for example, in medical diagnosis) to take the risk. Yet we are also aware that such inference is ‘defeasible’ - that new information may undermine old conclusions. Various kinds of defeasible but remarkably successful inference have traditionally captured the attention of philosophers (theories of induction, Peirce’s theory of abduction, inference to the best explanation, and so on). More recently logicians have begun to approach the phenomenon from a formal point of view. The result is a large body of theories at the interface of philosophy, logic and artificial intelligence.

1 Sources of non-monotonicity

A relation of inference is ‘monotonic’ if the addition of premises does not undermine previously reached conclusions; otherwise the relation is non-monotonic. Theories of non-monotonic inference locate the source of non-monotonicity in various forms of calculated risk-taking in the course of seeking information. Many kinds of reasoning involving ‘degrees of belief’ provide familiar examples. Suppose that a set A of assumptions raises the credibility of an event described by a sentence b above a certain threshold. (Lower case letters will stand for sentences in a given language, upper case letters denote sets of sentences.) Then it is quite possible that the credibility of b falls below that threshold if further assumptions are added to A . Thus, while one may be prepared to accept b on the assumptions A , one may prefer to remain agnostic about b given additional information. Credibility here may be measured by ordinary probability or by more general measures, such as Shackle’s (1979) non-additive degrees of belief (where the credence of mutually exclusive and jointly exhaustive alternatives need not add up to unity). Indeed, most advocates of non-monotonic formalisms insist that they model a wider class of phenomena than can be captured by ordinary probability functions.

A particularly simple source of non-monotonicity is the ‘closed world assumption’ (CWA). With respect to a given collection of data about some subject matter it states that the collection is complete; that no relevant information is left out. Hence, if a is not part of the collection, it is safe to assume $\neg a$. Since a may be derivable from a larger set of data, derivability under the CWA is non-monotonic. The CWA is behind the treatment of ‘negation as failure’ in Prolog (‘Programming in logic’, a programming language): Prolog derives $\neg a$ from a set of data D just in case a query for a given D fails. (Caution: a must be of a suitably simple syntactic kind. For complex sentences, negation as failure is a delicate and potentially hazardous matter.)

Another source of non-monotonicity is the licence to draw from a pool of presumptions (‘defaults’) as long as these are ‘safe’ to use (that is, consistent). For example, if one is told that tea is served, one will naturally act on the presumption that it is served plain or with milk and/or sugar. That presumption may become unsafe as additional information is supplied (such as that tea is served with brake fluid or salt).

Default logic (Reiter 1980) is one of the first and most influential theories responding to this source of non-monotonicity. It serves well as a principal example of a non-monotonic logic because it exhibits very clearly three characteristic features of almost all theories of non-monotonic reasoning: first, closure under rules with premises that require tests for non-derivability (‘non-Horn’ rules); second, a non-unique process of extending premises using defaults; and third, some method for reducing the multitude of extensions and for specifying a univocal set of conclusions.

A ‘default theory’ (D, A) (giving rise to a default logic) consists of a set A of premises (or ‘facts’) and a set D of default rules (‘defaults’ for short), each of the form

$$(1) \quad (B : C) \Rightarrow a.$$

These rules are to be read as follows: infer a from B unless C is inconsistent with what can be known by default on

the basis of A .

A similar and, under certain conditions, equivalent way of expressing defaults is by modal formulas such as

$$(1') \quad \Box b \wedge \neg \Box \neg c \rightarrow a,$$

as in Moore's (1985) auto-epistemic logic, where $\Box a$ may be read ' a is known'.

The rule (1) is non-Horn, since the 'unless' clause must be verified by showing that no contradiction can be derived from C together with the closure of A under the default rules.

Closing the premises A under the default rules in D issues in an extension of the default theory (D, A) . More precisely, let $\Sigma = (S_0, S_1, \dots)$ be a sequence of sets of formulas with $S_0 = A$, let S be the union of all its members, and let each S_i in Σ be such that

$$S_{i+1} = \{a : \text{there is a default } (B : C) \Rightarrow a \text{ in } D, \\ B \subseteq Cn(S_i) \text{ and} \\ C \text{ is consistent with } S\}$$

(where Cn denotes the operation of generating the deductive consequence of a given set). For each such sequence Σ , $Cn(S)$ is called an 'extension' of the default theory (D, A) .

To see how a single default theory (D, A) gives rise to a multitude of extensions, let $A = \{a\}$ and consider a set D of two defaults,

$$(2) \quad (a : \neg b) \Rightarrow c$$

$$(3) \quad (a : \neg c) \Rightarrow b$$

If we first close A under (2), we obtain a set A_1 containing c , so b cannot be inferred using (3). If we first close A under (3) we obtain a set A_2 containing b , so c cannot be inferred using (2). Thus, the result of extending a set of premises ('facts') by using default rules depends on the order in which the rules are applied. In general there are many extensions of a set of facts under default rules; sometimes there are none (Poole 1994: §7).

The multiplicity can be resolved by intersecting a set of selected extensions. There are three basic options for selecting: according to the 'sceptical' (full meet) solution, all extensions should be selected; the 'brave' (maxi-choice) approach recommends selecting a single extension; and the 'cautious' (partial meet) approach is less committal in requiring only that some extensions be selected. In each case a set of premises A is said to entail a conclusion b by default D just in case b follows deductively from the intersection of all selected extensions of A under D . How should one select among possible extensions? Some approaches recommend that the basis should be syntactic information about the applied defaults: for example, that extensions generated by more specific defaults should take precedence over those generated by more high-handed ones. Others assume that defaults are to be ordered, either by a qualitative preference ranking or by a numerical assignment of degrees of belief.

2 Patterns and models of non-monotonic inference

Inference is 'defeasible' if new information may undermine old conclusions. Formal theories of non-monotonic reasoning differ greatly both as to the kind of defeasibility they focus on and as to their chosen level of generality. They also differ, at first sight, from what are typically referred to as 'logics'. Still, the term logic is appropriate because all these theories can be used to specify a relation in some formal language which exhibits many of the patterns usually associated with relations of inference.

It is not enough to characterize defeasible inference negatively by the possible failure of monotonicity. We also need to ask: which conditions are compatible with defeasibility and which conditions should be satisfied by any relation of inference? These questions point towards continuing the abstract study of consequence relations (or operations) initiated by Tarski (1930). By abstracting from the particular ways in which inference relations are generated we turn to properties they may share.

Let $Cn(A)$ stand for the set of classical and $C(A)$ for the set of defeasible consequences of some set A . Among the more stable properties for C that have emerged are the following:

to $A \subseteq C(A)$ (inclusion) $C(A) = C(C(A))$ (idempotence) If $A \subseteq B \subseteq C(A)$ then $C(A) = C(B)$ (cumulativity) $C_n(A) \subseteq C(A)$ (supraclassicality) $C(A) \cap C(B) \subseteq C(C_n(A) \cap C_n(B))$ (distribution)

In the presence of the other conditions, cumulativity is equivalent to the slightly more perspicuous: from $A \subseteq C(B)$ infer $C(A \cup B) = C(A)$. The properties of C may be recast in terms of a relation $A \sim b$, defined to mean $b \in C(A)$. For example, cumulativity combines the classical principles of cut,

If $(\forall b \in B) A \sim b$ and $A \cup B \sim c$, then $A \sim c$ (cut)

with a cautious form of monotonicity,

If $(\forall b \in B) A \sim b$ and $A \sim c$, then $A \cup B \sim c$. (cautious monotonicity)

Inference operations (and their associated relations) satisfying the first three conditions are called ‘cumulative’; if they also satisfy the last two conditions, they are called ‘classically cumulative’. Note that although the above list describes the common core of various theories of non-monotonic inference, theories based on high probability rules for acceptance are not among them. If $A \sim b$ is defined to hold just in case the probability of b given all members of A reaches a certain threshold, then $\sim$ satisfies neither cut nor cautious monotonicity, nor can $\sim$ be both idempotent and supraclassical; for that would entail the objectionable left absorption principle $C = C_n C$. (Objectionable because of ‘lottery paradoxes’ where both a and b are accepted, while their logical consequence $a \wedge b$ is not (see Kyburg 1961).) Still, high probability accounts of $\sim$ can be made to score better on the above properties, if the probability required for acceptance is pushed high enough, that is, infinitely close to unity. Such accounts have been developed by Adams (1975) and Pearl (1988).

Cumulative inference can be characterized in terms of *preferential models* (see Shoham 1987; Kraus, Lehmann and Magidor 1990; Makinson 1994). These consist of a set of ‘states’ ordered by some relation $<$ and a mapping $| \cdot |$ assigning to each sentence a set of states. Intuitively, the mapping takes a sentence to a state just in case the sentence truly describes the state, and the relation $<$ orders states with respect to their ‘extravagance’ (conversely, ‘naturalness’). For example, a state in which tea is served with a dash of brake fluid is less natural (more extravagant) than a state in which tea is served with salt, which in turn is still less natural than serving tea with sugar or with milk or with both. The most natural states are the states we can expect to obtain unless there is evidence to the contrary.

If $|a|$ is the set of all states at which a holds, then the minimal elements of $|a|$ under $<$ represent the most natural (‘preferred’) ways for a to be true. The condition of ‘stopperedness’ (sometimes called ‘smoothness’; see also the ‘limit’ assumption in Lewis 1973) requires that there always are minimal elements.

For each set A of sentences, let $|A|_<$ denote the set of minimal states at which all sentences in A hold. Then A *preferentially* entails b just in case $|A|_< \subseteq |b|$. In contrast to ordinary entailment, preferential entailment restricts the domain of quantification: not all possible but only the most natural ways in which all elements of A can be true are considered. Since a most natural way for A to be true may not be a most natural way for $A \cup B$ to be true, preferential entailment is not monotonic. To illustrate: we may find that all most natural ways of making the sentence ‘tea is served’ true will please someone in the room. But that finding may well be undermined, if we move to considering all ‘natural’ ways of serving tea with brake fluid.

It can be shown that (stoppered) preferential entailment is a cumulative inference relation and that every cumulative inference relation can be presented as a relation of preferential entailment. (For the representation results mentioned here see Makinson 1994: §3.4.) The condition of stopperedness guarantees that preferential entailment will be cautiously monotonic. (To see how cautious monotonicity may fail in non-stoppered models let the states be indexed by the natural numbers and let s_m be more natural than s_n just in case m is greater than n . Consider three sentences a, b, c such that a holds everywhere, c fails everywhere and b holds at the first state only. Then $|a|_< = \emptyset$ because there is no most natural state for a . So, on the one hand we have both $|a|_< \subseteq |b|$ and $|a|_< \subseteq |c|$ vacuously. On the other hand, there is exactly one state - namely the first - that satisfies both a and b . Hence, $|\{a, b\}|_<$ - having the first state as its (only) member - cannot be contained in $|c|$ since the latter was assumed to be empty.)

If the mapping $| \cdot |$ is classically well-behaved (that is, satisfies the equations $|a \wedge b| = |a| \cap |b|$ and $|\neg a| = \overline{|a|}$), then each state will be closed under classical consequence whence preferential entailment will be supraclassical. Moreover, it will also be distributive and thus satisfy the conditions for being classically cumulative. Conversely,

each classically cumulative inference can be represented as preferential entailment in preferential models that respect the truth-functional connectives.

Supraclassicality is a rather common property of non-monotonic reasoning. It gives formal expression to the idea that such reasoning goes beyond what can be inferred ‘safely’ (deductively) from a given body of knowledge. However, some theories of non-monotonic inference (like those based on defeasible inheritance nets or on reason maintenance systems) fail to fully extend classical logic because they operate in impoverished fragments of classical languages. Such restrictions are usually introduced for computational reasons; they are not based on principled objections to supraclassicality.

Note also that in so far as non-monotonic logics are supraclassical, they are not rivals but extensions of classical logic. This sets them apart from certain other logics which fail to be monotonic, such as relevance, linear and many paraconsistent logics. These logics are *subclassical* and their source of non-monotonicity is different, having nothing to do with ‘risk-taking’.

3 Related work

A sentence *b* may be (defeasibly) inferred from a set *A* in the light of background assumptions *B* just in case *b* follows logically from *A* together with as much of *B* as is compatible with *A*. The right-hand side of this equivalence can be seen as involving a step of belief change: from the background beliefs *B* to a *revision* of *B* so as to accommodate *A* (with a minimal amount of change). Belief revision and defeasible inference thus appear to be two sides of the same coin.

The correspondence between belief revision and non-monotonic reasoning has been investigated in depth by Gärdenfors and Makinson (1994). In contrast, Levi (1995) argues that the correspondence can only be partial because the sole purpose of belief revision is to accommodate belief-contravening evidence. Such accommodation obeys a maxim of minimal change. But in so far as non-monotonic reasoning also covers ampliative inference, it must go beyond the maxim of minimal change governing belief revision. Hence, according to Levi, properties of non-monotonic inference as generated by belief revision on the one hand and as generated by ampliative inference on the other hand cannot be expected to coincide.

There is a close relation between counterfactual conditionals and belief revision via the ‘Ramsey test’: accept a conditional ‘If *a* were the case, then *b* would be the case’ in a state of belief *A* just in case you would accept *b* if *A* had to be revised so as to contain *a*. Given the correspondence, noted in the last paragraph, between belief revision and non-monotonic reasoning, one expects - and will not be disappointed - to find similarities between counterfactuals and non-monotonic reasoning.

A more direct route to such similarities is opened by comparing preferential models with the Stalnaker-Lewis semantics for counterfactuals (see [Counterfactual conditionals §§3-4](#)). In this semantics a counterfactual is pronounced true in a ‘world’ *w* just in case the consequent is true in all the closest possible worlds (that is, those worlds which depart as little as possible from *w*) that make the antecedent true. There is an obvious similarity between the condition for evaluating preferential entailments and that for evaluating counterfactuals: not all antecedent-worlds are examined but only those that are minimal under a certain relation. But there are also differences. Because each model for counterfactuals is equipped with as many relations of closeness as there are worlds to be close to, counterfactuals can be evaluated (differently) at single worlds. By contrast, preferential models have only one relation; accordingly, preferential entailments are assessed not with respect to single worlds but only in a model as a whole. It is a controversial matter how to interpret such subtle differences (see [Makinson 1993](#)): whether preferential entailments are essentially valid counterfactuals frozen at the first degree (that is, with the conditional connective joining truth-functional formulas only) or whether the differences are more substantial.

See also: [Logical and mathematical terms, glossary of](#)

ANDRÉ FUHRMANN

References and further reading

- Adams, E.** (1975) *The Logic of Conditionals*, Dordrecht: Reidel. (Derives a relation of defeasible inference (*p*-entailment) from limiting probability valuations.)
- Brewka, G.** (1991) *Non-Monotonic Reasoning: Logical Foundations of Commonsense*, Cambridge: Cambridge

University Press.(A concise introduction to the topic.)

- Gabbay, D.M., Hogger, C.J. and Robinson, J.A. (eds)** (1994) *Handbook of Logic in Artificial Intelligence and Logic Programming*, vol. 3, *Non-Monotonic Reasoning*, Oxford: Oxford University Press.(A collection of authoritative surveys.)
- Gärdenfors, P. and Makinson, D.** (1994) 'Non-Monotonic Inference Based on Expectations', *Artificial Intelligence* 65: 197-245.(Investigates correspondences between belief change and non-monotonic reasoning.)
- Kraus, S., Lehmann, D. and Magidor, M.** (1990) 'Non-Monotonic Reasoning, Preferential Models and Cumulative Logics', *Artificial Intelligence* 44: 167-207.(A study of models for non-monotonic inference relations extending the approach of Shoham (1987).)
- Kyburg, H.** (1961) *Probability and the Logic of Rational Belief*, Middleton, CT: Wesleyan University Press.(An early treatise of defeasible inference as in high probability accounts of acceptance. See also Kyburg's contribution to Gabbay et al. (1994).)
- Levi, I.** (1995) *For the Sake of the Argument: Ramsey Test Conditionals, Inductive Inference and Non-monotonic Reasoning*, Cambridge: Cambridge University Press.(Critically examines non-monotonic logic in the light of philosophical studies of inductive inference and rational belief change.)
- Lewis, D.K.** (1973) *Counterfactuals*, Oxford: Blackwell.(Seminal study of models for counterfactual conditionals.)
- Makinson, D.** (1993) 'Five Faces of Minimality', *Studia Logica* 52: 339-79.(Compares various ways of using the notion of 'minimality' in semantics for non-monotonic logic, semantics for conditionals and for conditional obligation, belief change and updating.)
- Makinson, D.** (1994) 'General Patterns in Non-Monotonic Reasoning', in D.M. Gabbay, C.J. Hogger and J.A. Robinson (eds) *Handbook of Logic in Artificial Intelligence and Logic Programming*, vol. 3, *Non-Monotonic Reasoning*, Oxford: Oxford University Press.(Focuses on the material of §2.)
- Moore, R.C.** (1985) 'Semantical Considerations on Non-Monotonic Logic', *Artificial Intelligence* 25: 75-94.(Develops the modal version of default logic.)
- Pearl, J.** (1988) *Probabilistic Reasoning in Intelligent Systems*, San Mateo, CA: Morgan Kaufmann.(Derives non-monotonic inference from limiting probability valuations. Rediscovered and further develops Adams' (1975) theory of p -entailment (under the name of ε -entailment).)
- Poole, D.** (1994) 'Default Logic', in D.M. Gabbay, C.J. Hogger and J.A. Robinson (eds) *Handbook of Logic in Artificial Intelligence and Logic Programming*, vol. 3, *Non-Monotonic Reasoning*, Oxford: Oxford University Press.(A survey of default logics.)
- Reiter, R.** (1980) 'A Logic for Default Reasoning', *Artificial Intelligence* 13: 81-132.(Classic presentation of default logic.)
- Shackle, G.L.S.** (1979) *Imagination and the Nature of Choice*, Edinburgh: Edinburgh University Press.(Shackle's writings have informally anticipated many of the models of non-monotonic reasoning.)
- Shoham, Y.** (1987) 'A Semantical Approach to Non-Monotonic Logics', in *Readings on Non-Monotonic Reasoning*, ed. M. Ginsberg, Los Altos, CA: Morgan Kaufmann.(First presentation of preferential models.)
- Tarski, A.** (1930) 'Fundamentale Begriffe der Methodologie der deduktiven Wissenschaften I', *Monatshefte für Mathematik und Physik* 37: 361-404; trans. J.H. Woodger (1956), 'Fundamental Concepts of the Methodology of Deductive Sciences', in *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*, ed. J. Corcoran, Indianapolis, IN: Hackett Publishing Company, 2nd edn, 1983, 60-109.(Classic study of abstract consequence operations.)
- Ullman-Margalit, E.** (1973) 'On Presumptions', *Journal of Philosophy* 80: 143-63.(Develops informally many of the ideas underlying formal theories of non-probabilistic non-monotonic reasoning.)

Numbers

Numbers are, in general, mathematical entities whose function is to express the size, order or magnitude of something or other. Historically, starting from the most basic kind of number, the positive integers (1, 2, 3,...), which appear in the earliest written records, the notion of number has been generalized and extended in several different directions - often in the face of considerable opposition.

Other than the positive integers, the most venerable are the rational numbers (fractions), which were known to the Egyptians and Mesopotamians. The discovery, by Pythagorean mathematicians, that there are lengths that cannot be expressed as fractions occasioned the introduction of irrational numbers, such as the square root of 2, though the Greeks managed only a geometric understanding of these. The number zero was recognized, first in Indian mathematics, by the seventh century; the use of negative numbers evolved after this time; and complex numbers, such as the square root of -1 , appeared first at the end of the Middle Ages. Infinitesimal numbers were developed by the founders of the calculus, Newton and Leibniz, in the seventeenth century (and were later to disappear from mathematics - for a time); and infinite numbers (ordinals and cardinals) were introduced by the founder of modern set theory, Cantor, in the nineteenth century.

The introductions of three of these kinds of number, in particular, occasioned crises in the foundations of mathematics. The first (concerning irrational numbers) was finally resolved in the nineteenth century by the work of Cauchy and Weierstrass. The second (concerning infinitesimals) was also resolved then, by the work of Weierstrass and Dedekind. The third (concerning infinite numbers), which involves paradoxes such as Russell's, still awaits a convincing solution.

It is seemingly impossible to give a rigorous definition of what it is to be a number. The closest one can get is a family-resemblance notion, with very ill-defined boundaries.

1 Natural numbers

All human societies would appear to have some form of counting, however limited, both for size - cardinality (one, two, three,...) - and for order - ordinality (first, second, third,...). The recognition that an infinitude of positive integers could be used for these purposes was already in place in the earliest civilization from which we have written records, that of Egypt *circa* the fourth millennium BC. The investigation of basic arithmetic operations (addition, multiplication and so on) was also well under way then. This was continued in the next great Middle Eastern civilization, that of Mesopotamia (Babylon), *circa* the second millennium BC. The investigations were greatly facilitated by the Mesopotamian invention of place-notation, that is, the idea that a single numeral can represent different quantities depending on where it occurs in a string of numerals. Thus, for example, in decimal notation, the first '2' of '2,121' represents two thousands and the second represents two tens. (The Mesopotamians actually preferred a sexagesimal to a decimal base.)

The *natural* numbers comprise the positive integers together with zero. The idea of zero as a number was not to be found in either of these civilizations, although the Mesopotamians did sometimes leave a gap where we would now write a zero. (Thus '2 2' might represent 202 or 2,020, and so on.) Though the origin of the idea is uncertain, zero was certainly being used by the seventh century AD by Indian mathematicians, who had also developed place-notation and used a decimal base. They therefore possessed, in effect, our modern number system (though the symbols they used were different). This system was taken up in the Arabian civilization around the eighth century, and thence passed into Europe around the twelfth.

2 Rational and irrational numbers

Positive integers quantify wholes. Once one starts to measure, it becomes clear that a way is required to quantify parts. *Rational* numbers, as ratios of natural numbers (with non-zero denominator), clearly serve this function. These were known to the Egyptians, at least in the form of unit fractions (of the form $1/n$) and, in a more sophisticated and general form, to the Mesopotamians. Decimal fractions (of the form 0.23) appear to have been developed by Arabic mathematicians, and by the time we reach Stevin in late sixteenth-century Europe, the modern computational system for rational numbers is essentially in place.

The creation of modern mathematics is usually reckoned to have occurred in Ancient Greece in the second half of the first millennium BC. What was distinctive about the Greek approach was that mathematics was freed from its practical roots and took on a purely theoretical form. In particular, the central concern of Greek mathematics was proof, especially proofs in geometry. As far as arithmetic goes, the first part of this epoch was dominated by Pythagoreanism, according to which everything can be explained in terms of the positive integers. (It follows that rational numbers are of little theoretical importance: given two lengths to compare, if one chooses one's unit of measurement small enough, the lengths will be integral.)

The discovery that the Pythagorean assumption was wrong came as a distinct shock. Most simply, the diagonal of a unit square (whose length is the square root of two) is not commensurable with its sides. The proof of this (one of the first *reductio* arguments in the history of mathematics) is now celebrated. It is not known who discovered it, though it was probably one of the later Pythagoreans, c.400 BC. At any rate, it inaugurated the first of three crises in the foundations of mathematics that have been associated with the introduction of a new kind of number.

The new numbers in this case are *irrational* numbers (that is, numbers which cannot be written as a ratio of integers - written as decimals they are infinite and non-repeating), such as the square root of two (and π), although the irrationality of this was not proved until the eighteenth century by Lambert). They are required, together with rational numbers, for general quantification of length: in fact, the only conception of such magnitudes available to the Greeks was as geometric lengths. Given this representation, some geometric account of arithmetic operations had to be given. The hardest of these is division, which was solved by Eudoxus (c.360 BC) in his theory of proportions, given in book 5 of Euclid's *Elements*. But even for Eudoxus, ratios were not entities in their own right. Operations which treat irrational numbers in the form of surds (expressed in ways such as $\sqrt{2}$, $\sqrt[3]{71}$), as bona fide entities, can be found in Indian mathematics and, in particular, in the seventh-century mathematician Brahmagupta. Some use of surds was made by Arabic mathematicians, but their use did not become common until the sixteenth century (for example, in the work of Stevin and Cardan). A clear statement to the effect that irrational numbers are first-class numerical citizens can be found in Newton, in the seventeenth century; however, an adequate understanding of what this citizenship amounted to had to wait until the nineteenth century.

3 Negative and complex numbers

Subtraction had been familiar to mathematicians since the earliest times. The idea that a sensible quantity might itself be negative, however - and what the sense of it might be - took a long time to catch on. *Negative* numbers, the rules for operating with them, and their use as roots of equations appear in Brahmagupta, and were consequently known to Arabic mathematicians - though they did not use them as roots of equations. The first time they appear in this way in European mathematics is in the work of fifteenth-century mathematician Nicholas of Chuquet (who, however, did not allow zero to be a root!). Even Cardan, a century later, who employed negative roots extensively, called them 'fictional'. The idea that negative numbers might correspond to direction reversal is to be found in the early seventeenth-century mathematician, Girard; but, even after the invention of analytic geometry by Descartes a little later, many mathematicians simply ignored the negative parts of the plane. And even as late as the middle of the eighteenth century some textbook writers were still disputing the claim that the product of a pair of negative numbers is positive.

Once the use of negative numbers became common mathematical practice, the thought that there might be a new kind of number to provide for their square roots was not far behind. If we call all of the numbers mentioned so far *real* numbers, and write the square root of -1 as i (following Euler), the product of i with any real number is called an *imaginary* number. More generally, the sum of a real number and an imaginary number ($a + ib$) is called a *complex* number. One of the earliest occurrences of complex numbers in mathematics is in the work of Cardan, who observed, as we would now put it, that a cubic equation with a real root may also have complex roots, though Cardan himself regarded this observation as useless. A few years later, Bombelli articulated the use of complex numbers (in the same context) much further. Opposition to them did not then cease, however; for example, their use was frowned upon by Newton. But by the time of, and especially in the work of, Euler, in the eighteenth century, complex numbers had been shown to have numerous uses, both as the solutions of equations and elsewhere - for example, in important relationships between trigonometric functions. And by the time that Gauss, in the early nineteenth century, showed that every polynomial equation with complex coefficients has complex roots, complex numbers were well entrenched. With the development of electromagnetic theory later in the

century, complex numbers were shown to be capable of representing even physical quantities (for example, impedance).

The problem of what sense to make of complex numbers was also solved by Gauss and later contemporaries, such as Argand. The complex number $x + iy$ could be thought of as the point (x, y) in two-dimensional Euclidean space (now usually called the Argand plane), with the arithmetic operations on complex numbers defined in a suitable way. This prompted the idea that points in a higher-dimensional space could also be thought of as numbers of a certain kind. Defining a suitable notion of multiplication for points in a three-dimensional space turned out to be impossible, but the problem was solved for four dimensions by the Irish mathematician William Hamilton later in the nineteenth century. Hamilton defined a class of numbers known as ‘quaternions’, of the form $x + iy + jz + kw$, where $ijk = i^2 = j^2 = k^2 = -1$. The properties of quaternions were investigated for some years, but no real use was found for them, so the investigations lapsed. By this time, the observation that a somewhat different set of operations on points in Euclidean space might be fruitful had been made, and vector algebra was born. (If one thinks of complex numbers as two-dimensional vectors, then complex addition and vector addition are the same thing, but complex multiplication and the various vector multiplications are distinct.)

4 Infinitesimal numbers

The numbers discussed so far have all been finite. But numbers of some kinds are infinite - or their intuitive inverse; infinitesimal. It is sometimes (dubiously) claimed that infinitesimals can be found in Greek mathematics. What certainly can be found, from about the third century BC on, is a method called ‘the method of exhaustion’, (probably) invented by Eudoxus and developed by Archimedes. The main purpose of this method is the computation of areas and volumes of non-rectilinear geometric figures, and requires their approximation by rectilinear figures. For example, a circle is approximated by a regular polygon inscribed within it. As the length of the sides of the polygon decreases (and the number of sides increases) the approximation gets closer. By the late sixteenth century we find mathematicians such as Stevin and Wallis suggesting, in effect, that if we make the sides of the polygon infinitesimally small, the approximation will be exact - or, at least, will differ only infinitesimally from the true value. Hence, a new kind of number, *infinitesimals*, had to be recognized, to quantify infinitesimal lengths.

The use of infinitesimals in essentially this way was developed by Newton and Leibniz in the next century as the foundation of modern calculus (see [Analysis, philosophical issues in §1](#)). Though their inventions were independent, and used quite different notations, the essential ideas were the same. For example, to determine the gradient of the curve $y = x^2$ at the point $x = a$, we consider a point on the curve an infinitesimal distance d away along the x -axis. The slope of the line joining these two points is

$$\frac{(a+d)^2 - a^2}{d} = \frac{2ad + d^2}{d} \\ = 2a + d$$

but since d is infinitesimally small, we can disregard it. Hence the gradient is $2a$.

The power of infinitesimal methods was so great, in both pure and applied mathematics, that they soon became entrenched. However, many people - most notably Berkeley - were severely critical. In particular, infinitesimals had at the same time to be both non-zero (since one divided by them) and zero (since one ignored them). This dilemma inaugurated the second foundational crisis in mathematics associated with the introduction of a new kind of number. But this one would have to wait only 200 years for a solution.

5 Transfinite numbers

It is clear that there are infinite quantities in mathematics, and hence that there must be a size that these quantities can be, namely, infinite. But for most of the history of mathematics, infinitude was thought of as mathematically indeterminate, and hence there was little to be said about the quantity ‘infinity’. (One does, however, sometimes find mathematicians - even of the stature of Euler - taking it as a definite quantity, in particular, as the reciprocal of zero.) Even in the heyday of the infinitesimal calculus, infinite numbers played little part (though Leibniz certainly held that the reciprocals of distinct infinitesimals were distinct infinite numbers). The recognition of different orders of infinity, and consequently of the sense of infinite numbers, therefore came as another highly contentious

idea.

This recognition is almost entirely due to Georg Cantor in the late nineteenth century. He realized that, given some iterable operation (for example, forming the set of topological limit points of some set of points), it made sense to consider the result of having applied the operation an infinite number of times, and then to apply the operation again, and again...indefinitely (see [Cantor, G. §1](#)). To quantify this ordering, Cantor introduced a new class of infinite numbers, *transfinite ordinals*, the least of which he denoted by a lower-case omega, ω (see [Set theory §2](#)). He also defined arithmetic operations on these numbers which generalize the operations on finite numbers, but which have rather unusual properties. For example, addition is not commutative: $\omega + 1 \neq 1 + \omega$.

Having generalized the notion of order into the infinite, Cantor next generalized the notion of size. To do this he needed a criterion of size that would work (even) in infinite domains. He adopted the criterion (also suggested by [Frege \(§8\)](#)) that two collections have the same size just if their members can be put into one-to-one correspondence (that is, each member of either collection corresponds to exactly one member of the other). He observed that adding members to an infinite set may not increase its size. (This fact had been noted by some previous mathematicians - for example, Galileo - who thought it so absurd that they rejected infinity as a quantity.) In another of the most famous *reductio* arguments in the history of mathematics, Cantor also established that for any infinite collection, there is one of greater size. This is now called 'Cantor's theorem' (see [Cantor's theorem](#)). To quantify the different sizes Cantor introduced a new kind of infinite number, *transfinite cardinals* (see [Set theory §3](#)), denoted by the Hebrew letter aleph ($\aleph$) with a subscript ($\aleph_0$ being the smallest, the size of the natural numbers); and defined arithmetic operations on these which, again, generalize the finite case. These operations also have striking properties: for example, double any aleph and you get what you started with.

It should be noted that natural numbers can be thought of indifferently as ordinals or cardinals. This is because, canonically, we count the size of a finite collection by ordering, and its cardinality is n just if the last object in the ordering is the n th. However, transfinite ordinals and cardinals are quite distinct since, as just observed, if one adds further objects to an ordered set, the size may remain the same. Transfinite ordinal and cardinal arithmetics are also quite distinct. Cardinal addition, for example, is commutative, where ordinal addition is not. It is worth noting that there are very simple questions concerning basic transfinite arithmetic operations that are still unanswered (see [Continuum hypothesis](#)).

Cantor's introduction of transfinite numbers into mathematics did not entail the use of infinitesimals. This is because the unusual properties of (cardinal and ordinal) multiplication allow no obvious sense to be given to the notion of division, and so not to reciprocation. However, the use of infinite numbers brought its own problems. In particular, the totality of all ordinal numbers is an ordered collection and hence must have an ordinal, which must be greater than, and so distinct from, all ordinals. Similarly, the totality of all objects must have the largest possible cardinal, but, by Cantor's theorem, there is a larger. These paradoxes are now called the Burali-Forti paradox and Cantor's paradox, respectively; Russell's paradox, concerning the set of all sets that are not members of themselves, is a stripped-down version of Cantor's (see [Paradoxes of set and property §4](#)). They heralded the third crisis in the foundations of mathematics associated with the introduction of a new kind of number; a crisis which, unlike the first two, is as yet not satisfactorily resolved. These paradoxes fuelled the rejection of Cantor's ideas by some, including Kronecker; but despite this, they were absorbed into orthodox mathematics within about fifty years.

6 Weierstrass and Dedekind

The final resolution of the first two crises we met (concerning irrationals and infinitesimals) took place in the nineteenth century, a period when mathematicians became particularly concerned with the rigorous foundations of their subject. In this context, the work of Weierstrass and [Dedekind](#) is particularly significant.

Early in the nineteenth century, the notion of a 'limit' appeared in Cauchy's formulation of the calculus. His method differed from that set out in §4 above as follows: instead of taking d to be some infinitesimal quantity, we let it be a finite quantity, and then consider the limit of what happens when d approaches zero (the limit being a quantity that may be approached as close as we please, though never, perhaps, attained). Despite the fact that Cauchy possessed the notion of a limit, he mixed both infinitesimal and limit terminology, and it was left to Weierstrass, later in the century, to replace all appeals to infinitesimals by appeals to limits. At this point infinitesimal numbers disappeared from mathematics (though they would return, as we shall see).

Weierstrass also gave the first modern account of negative numbers, defining them as signed reals, that is, pairs whose first members are reals and whose second members are ‘sign bits’ (‘+’ or ‘−’), subject to suitable operations.

A contemporary of Weierstrass, Tannery, gave the first modern account of rational numbers. An ‘equivalence relation’ is a relation R that is reflexive (xRx), symmetric (if xRy then yRx) and transitive (if xRy and yRz then xRz). Given an equivalence relation on a domain of objects, an ‘equivalence class’ is the set of all those things in the domain related to some fixed object. Tannery defined the rationals as equivalence classes of pairs of natural numbers (the second of which is non-zero), under the equivalence relation ‘ $\sim$ ’ (tilde), defined as follows.

$$\langle m, n \rangle \sim \langle r, s \rangle \text{ iff } m.s = r.n$$

The problem of irrational numbers was finally solved independently by Weierstrass, Cantor and Dedekind, who gave different but equivalent constructions of real numbers. Weierstrass’ is in terms of infinite decimal expansions; Cantor’s is in terms of convergent infinite sequences of rationals. Dedekind’s construction is the simplest: consider any splitting of the rational numbers (which can be taken to include the integers, since the integer n may be identified with the rational number $n/1$) into two non-empty disjoint collections L (‘left’) and R (‘right’), such that anything less than a member of L is in L , and anything greater than a member of R is in R . This pair is now called a ‘Dedekind cut’ (or ‘section’). Real numbers (including irrational numbers) can be thought of as such sections (or just one of their parts).

Dedekind also gave the first axiom system for the natural numbers, in terms of an initial number and the successor operation (+1). In modern form, the axioms are as follows:

- (1) 0 is a number.
- (2) The successor of any number is a number.
- (3) 0 is the successor of no number.
- (4) Any two numbers with the same successor are the same.
- (5) If 0 has some property, and the successor of any number with that property also has it, then all numbers have it.

These axioms, together with the recursive definitions for addition and multiplication (also given by Dedekind), are now usually named after Peano, who gave a formalized version a few years later.

7 Set-theoretic reduction

The drive for rigour in the foundations of number theory reached its height in the reduction of all numbers to sets, due to the logicians Frege, Russell and Whitehead at the turn of the twentieth century (see **Logicism**). The key to this was a set-theoretic definition of cardinal and ordinal numbers. The cardinal numbers were defined as equivalence classes of sets under the equivalence relation ‘can be put into one-to-one correspondence with’. (For Russell and Whitehead, this was restricted to sets of fixed type, in an attempt to avoid the paradoxes of infinite number - see **Theory of types**.) A ‘well-ordering’ on a set is an ordering such that every non-empty subset has a least member in the ordering (see **Set theory §2**). Two well-ordered sets are ‘order-isomorphic’ if they can be put into a one-to-one correspondence that preserves the ordering. The ordinal numbers were defined as the equivalence classes of well-ordered sets (of a given type) under the equivalence relation of order-isomorphism. Cardinal and ordinal arithmetic operations were defined in an appropriate fashion.

Given the ordinals, other numbers could then be defined in a relatively straightforward way. Natural numbers are the finite ordinals (which can be shown to satisfy the Peano axioms - see §6 above); rational numbers can be defined by the Tannery construction; real numbers as Dedekind sections; negative numbers as Weierstrassian pairs; complex numbers can be defined as pairs of signed reals, thought of as points on the Argand plane. In each case, arithmetic operations can be defined in natural ways.

Since the Frege/Russell reduction, several other non-equivalent, but equally good, definitions of the cardinal and ordinal numbers have been discovered (and in virtue of this, the claim that these numbers just *are* certain sets is difficult to maintain). The most elegant of these is due to von Neumann. According to this, each ordinal is simply the collection of all smaller ordinals. (Thus, zero is simply the empty set, the least infinite ordinal is the set of natural numbers, and so on.) Cardinal numbers are identified with ‘initial’ ordinals, that is, least ordinals of each

size.

8 Developments from logic

Work in mathematical logic in the twentieth century has provided several notable developments bearing on numbers. Three are particularly important. The first was the proof by Gödel in 1931 that the Peano axioms, and all other consistent axiom systems for arithmetic, are incomplete, in the sense that there are truths of arithmetic that cannot be proved from the axioms - at least if the underlying logic is first-order (see [Gödel's theorems](#)). The axioms are complete if the underlying logic is second-order and the induction principle (5) is formulated as a second-order axiom and not just a first-order schema; but second-order logic is not itself axiomatizable (see [Second- and higher-order logics §1](#)). This raises profound questions about the nature of both numbers and our knowledge thereof, that fall outside the bounds of this entry.

The second development concerns the paradoxes surrounding transfinite numbers. The orthodox view that has emerged this century is that embedded in Zermelo-Fraenkel set theory (ZF). According to this, there just is no totality of all ordinals, all sets or other 'large' collections, and so the question of their size does not arise. Although this account provides enough set theory for most mathematics (though not all: category theory appears to require large sets of just this kind), it can hardly be said to be conceptually adequate. For example, standard logic defines the sense of a quantifier in terms of the domain (totality) over which it ranges. It is therefore unclear what the sense of the quantifiers of ZF is, if, as it claims, there is no such totality. (See [Priest 1987](#).)

The third development is due to Robinson and is called 'nonstandard analysis' (see [Analysis, nonstandard](#)). As was proved originally by Löwenheim and Skolem, (first-order) theories of number have nonstandard models (see [Löwenheim-Skolem theorems and nonstandard models](#)). In particular, any theory of the reals will have such models. Robinson showed that in all of these models, there are non-zero numbers that are smaller than any real number: infinitesimals. Using these, he demonstrated that the reasoning of the infinitesimal calculus (which is much more intuitive than limit reasoning) can be interpreted in a perfectly consistent manner. Hence, infinitesimals have been rehabilitated as perfectly good numbers.

9 Number in general

The preceding review of the development of the notion of number naturally prompts the question of what a number is. One might interpret this as the question of whether numbers are Platonic objects, mental constructions, or nothing more than mystified numerals. This is a central issue in the philosophy of mathematics.

Alternatively, in virtue of the plethora of kinds of numbers we have seen, one might interpret the question as asking what makes entities of certain kinds, but not others, numbers. Beyond the rather vague characterization with which this article began, it seems difficult to give a general characterization of number. The most fundamental numbers (both historically and conceptually), the natural numbers, measured size (or order), were subject to distinctive operations (such as addition) and could be the roots of equations. Each of these central features has played a role in generating new kinds of numbers (different concerns being dominant on different occasions). The result is a collection of entities which are related by family resemblance (as observed by Wittgenstein in §67 of *Philosophical Investigations*), though the boundaries of the family seem somewhat arbitrary. It is difficult to see why, for example, complex numbers (or quaternions) should be called numbers, but not vectors or numerical matrices; both of these share the central features of natural numbers.

This conclusion is reinforced by recent work by Conway ([1976](#)). He gives a (transfinite) recursive construction that generalizes both the Dedekind construction of the reals and the von Neumann construction of ordinals. Essentially, a number is any pair, $\langle L, R \rangle$, such that all the members of L and R are numbers, and every member of R is greater than or equal to ($\geq$) every member of L (see [Analysis, philosophical issues in §2](#)). ' $\geq$ ' and the arithmetic operations are also defined in a natural recursive manner. The construction generates virtually all the numbers we have met in this article, including infinitesimals, but excluding, notably, the complex numbers (and if cardinals are to be identified with initial ordinals, a non-uniform definition of arithmetic operations is necessary). Moreover, the construction generates many novel numbers, for example, notably, numbers obtained by applying the full range of real-number operations to infinite numbers, for example, $\omega - 1$, $\sqrt{\omega}$, which make no sense on the usual understanding. Moreover, a simple generalization of the construction (dropping the ordering condition on

L and R), produces even more number-like objects (which Conway calls ‘games’, because, in a certain sense, they code the strategic possibilities in a two-person game).

Just conceivably, a unifying account of number might eventually be found, but in the meantime the emergence of new kinds of numbers seems likely. For example, there are nonstandard inconsistent models of arithmetic which contain inconsistent numbers (natural numbers with inconsistent properties). These have some notable applications. For example, some of them can be shown to provide solutions for arbitrary sets of simultaneous linear equations. (See [Mortensen 1995](#).) And just as the existence of nonstandard models of analysis made infinitesimals legitimate, so might these legitimize the notion of an inconsistent number.

See also: [Analysis, philosophical issues in](#); [Antirealism in the philosophy of mathematics](#); [Arithmetic, philosophical issues in](#); [Logical and mathematical terms, glossary of](#); [Realism in the philosophy of mathematics](#); [Set theory, different systems of](#)

GRAHAM PRIEST

References and further reading

- Bell, J.L. and Machover, M.** (1977) *A Course in Mathematical Logic*, Amsterdam: North Holland. (Chapter 10 gives a clear account of the modern set-theoretic reduction of cardinal and ordinal numbers.)
- Benacerraf, P.** (1965) ‘What Numbers Could Not Be’, *Philosophical Review* 74: 47-73; repr. in P. Benacerraf and H. Putnam (eds) *Philosophy of Mathematics: Selected Readings*, Cambridge: Cambridge University Press, 2nd edn, 1983. (A standard critique of set-theoretic reductionism.)
- Boyer, C.B. and Merzbach, U.C.** (1968) *A History of Mathematics*, New York and London: Wiley & Sons; 2nd edn, 1989, repr. 1991. (An excellent reference on the history of mathematics, recently updated.)
- Cajori, F.** (1919) *A History of Mathematics*, London: Macmillan; 5th edn, New York: Chelsea, 1991. (A classic on the history of mathematics. Now a bit dated, but still good value.)
- Conway, J.H.** (1976) *On Numbers and Games*, London: Academic Press. (Provides a unified construction of many different kinds of number. Cited in §9.)
- Crossley, J.N.** (1980) *The Emergence of Number*, Singapore: World Scientific. (A very readable history of number.)
- Dedekind, R.** (1963) *Essays on the Theory of Numbers*, New York: Dover. (Includes English translations of two of the most important essays on numbers ever written.)
- Lakatos, I.** (1978) ‘Cauchy and the Continuum’, in *Collected Papers*, vol. 2, *Mathematics, Science and Epistemology*, Cambridge: Cambridge University Press, ch. 3. (A perceptive analysis of the infinitesimal calculus, its downfall and its relationship to nonstandard analysis.)
- Mortensen, C.** (1995) *Inconsistent Mathematics*, Dordrecht: Kluwer. (Chapters 2 and 8 include discussions of inconsistent numbers and their use in solving inconsistent sets of linear equations. Cited in §9.)
- Nechaev, I.V.** (1987) ‘Number’, in M. Hazewinkel (ed.) *Encyclopaedia of Mathematics: An Updated and Annotated Translation of the Soviet Mathematical Encyclopaedia*, Dordrecht: Kluwer. (A clear but more technical account of number.)
- Priest, G.** (1987) *In Contradiction*, The Hague: Martinus Nijhoff. (Chapter 2 includes a discussion of problems of ZF. Cited in §8.)
- Wittgenstein, L.J.J.** (1953) *Philosophical Investigations*, ed. G.E.M. Anscombe and R. Rhees, trans. G.E.M. Anscombe, Oxford: Blackwell. (Numbers illustrate Wittgenstein’s notion of family resemblance. Cited in §9.)

Ontological commitment

A person may believe in the existence of God, or numbers or ghosts. Such beliefs may be asserted, perhaps in a theory. Assertions of the existence of specific entities or kinds of entities are the intuitive source of the notion of ontological commitment, for it is natural to think of a person who makes such an assertion as being 'committed' to an 'ontology' that includes such entities. So ontological commitment appears to be a relation that holds between persons or existence assertions (including theories), on the one hand, and specific entities or kinds of entities (or ontologies), on the other.

Ontological commitment is thus a very rich notion - one in which logical, metaphysical, linguistic and epistemic elements are intermingled. The main philosophical problem concerning commitment is whether there is a precise criterion for detecting commitments in accordance with intuition. It once seemed extremely important to find a criterion, for it promised to serve as a vital tool in the comparative assessment of theories. Many different criteria have been proposed and a variety of problems have beset these efforts. W.V. Quine has been the central figure in the discussion and we will consider two of his formulations below.

Many important philosophical topics are closely connected with ontological commitment. These include: the nature of theories and their interpretation; interpretations of quantification; the nature of kinds; the question of the existence of merely possible entities; extensionality and intensionality; the general question of the nature of modality; and the significance of Occam's razor.

1 Fundamental problems

The intuitive source of the notion of ontological commitment is people asserting things to exist. It may thus be thought of as a *binary relation* between persons or assertions of existence and the entities or kinds of entities they assert to exist. The main philosophical problem concerning commitment is whether there is a precise criterion for detecting commitments in accordance with intuition. Anyone who seeks a criterion of commitment must first decide what sorts of entities may properly stand in the positions of the relation. Unfortunately there are at least *five* significant candidates for occupying its first position: persons; acts of uttering (or inscribing); utterances (or inscriptions), that is, the actual 'tokens' produced; sentences uttered (or inscribed), that is, the 'types' of which tokens are produced; and propositions (which on some accounts are often expressed on occasions of utterance or inscription). And there is a similarly diverse range of candidates for the second relatum. So there is a daunting array of seemingly different ways to see commitment as a binary relation.

An encouraging thought about this proliferation is that it may not ultimately matter which approach one takes because, given any plausible choice, the relation may be thought of as holding *derivatively* between pairs of the unchosen sorts in a more or less natural way. It may also be that moving to a ternary (and so on) treatment would help absorb some of the complexity, but commitment is typically viewed as a binary relation, and we will follow that precedent here.

Another difficulty is that our intuitive talk of commitment is crucially ambiguous. Sometimes we speak as if a person who says there are whales is (thereby) committed to the existence of mammals, sometimes not. So, there is a sense of commitment that 'tracks entailments' and one that does not. Both senses are surely legitimate, but we should decide which we are discussing at the outset. If we distinguish thus between '*prima facie*' and 'all-in' commitment, then it is fair to say that the discussion provoked by W.V. Quine mainly concerns the former. This is not surprising given his famous antipathy to essential properties and analytic truth, as well as his explicit claims that the notion of commitment belongs to the 'theory of reference' (and not to the 'theory of meaning'). But we will see that some of Quine's formulations can seem more suitable for assessing all-in commitment than *prima facie*.

Quine's stated criteria are sometimes meant to reveal commitment to *specific* entities, sometimes to *kinds* of entities. We will focus on the latter. Here are two representative examples:

- (1) to say that a given existential quantification presupposes objects of a given kind is to say simply that the open sentence which follows the quantifier is true of some objects of that kind and none not of that kind.

(Quine 1953a: 131)

(2) entities of a given sort are assumed by a theory if and only if some of them must be counted among the values of the variables in order that the statements affirmed in the theory be true.

(Quine 1953b: 103)

There is a serious problem of interpretation for any criterion intended to apply to theories. (This includes (1), for Quine's mention of open sentences following quantifiers reveals that he is thinking of certain sorts of formal theory.) The problem is that we use 'theory' in at least three different ways. First, we attribute 'informal theories' to people on the basis of their natural-language claims. Then we 'represent' these informal theories formally. Here the canonical notion is that of an (axiomatic) first-order theory along with an 'interpretation', which consists of a non-empty domain (the 'range' of quantifiable variables) plus an assignment of 'extensions' based on that domain to the nonlogical symbols of the theory. Note that we have already spoken of the purely syntactic component as a 'theory', but we also regard the syntax plus the extensional semantics as an 'interpreted theory'. So an unqualified use of 'theory' could mean any of three different things.

Now it seems almost nonsensical to think of a purely syntactic object (such as an uninterpreted sentence or theory) as ontologically committed to anything (except perhaps the syntactic elements that comprise it). Unfortunately this is the most literal construal of (1). Another construal makes the interpreted theory (or sentence, and so on) bear the commitment. This makes much better sense, but it may not make for a much better criterion. What (1) does under this understanding is, in effect, to define a certain metalinguistic relation term, 'presupposes' (see [Presupposition](#)). The relation this term expresses is evidently ternary - it relates interpreted closed sentences, objects (from the domain) and (somehow) *kinds*. The definition utilizes the notions of a formula with one free variable, the satisfaction relation (understood as relating objects and formulas - Quine's 'true of'), objects and (somehow) *kinds*. So (1) offers a new piece of terminology in the metatheory of interpreted first-order theories, enhanced by whatever is needed to talk about kinds in whatever way is intended. To this extent it is harmless enough, but it is not clear what it has to do with intuitive commitment. Notice that on its basis we may distinguish two sorts of objects in the domain of a theory: those that are of kinds 'presupposed' by at least one existential sentence; and, as may occur, those that are not. But any theory with both sorts of entities is already committed to all of them in a far more basic way since they are all essential components of that theory. No theory without them is the same theory.

It may therefore seem best to view (1) as directed to the case of informal theories, and the same may be said of (2). The intuitive idea is perhaps that the 'ontology' of an interpreted formal theory is 'transparent': It simply consists of the members of its domain of interpretation (or a readily distinguishable subset of them). So one might hope to reduce the hazy question of the commitments of informal theories to the transparent case of their formal representatives. But if the commitments of a given informal theory are unclear, then it would seem that the choice of its formal representative must also be unclear. So it appears that any criterion conceived in this way cannot work: to apply it we must first find a representing theory, but we cannot do this unless we already know the informal commitments.

The search for a criterion to apply to *theories* may thus seem fated to produce either near nonsense, triviality or practical futility, depending on what is meant by 'theory'. This would not be welcomed by Quine, especially since the original idea was to find a neutral means for assessing something not always clearly in view. It was to help us compare theories, sometimes with something like Occam's razor in mind. (Of course there may be a criterion for informal theories that makes no similar appeal to formal representatives.)

2 Problems of detail

We now consider some problems that arise from the details of (1) and (2). A serious defect of (1) is that it offers no means of assessing informal theories that make false existence claims. Since there are no unicorns, there is no interpreted theory the domain of which contains unicorns, and so there is none with unicorns available to satisfy relevant open sentences.

(2) offers an improvement over (1) in just this respect. For it may be seen as ruling that there *must* be some unicorns in the domain of *any* interpreted theory that would be a suitable representative of the informal theory, and

there may be a way of understanding this that makes no intuitive commitment to unicorns.

Now let us return to the matter of whales and mammals. It is clear that if the informal claim that there are whales receives a suitable (interpreted) first-order formulation, then a relevant open sentence will be true of some mammals and not true of any non-mammals. So (1) finds the informal theory committed to mammals as well as to whales. Whether (2) has the same consequence evidently depends on whether whales *must* be mammals, not (as in (1)) on whether they actually are mammals. So a philosopher who accepts (2) and, like Quine, denies that any entities have any non-trivial essential (or necessary) properties may claim that the distinction between *prima facie* and all-in commitment is a distinction without a difference (see [Essentialism](#)).

But this means there is a further difficulty in trying to evaluate Quine's formulations in the style of (2): they yield different results depending on the background philosophical views of the person applying them. When Quine applies (2) to an informal theory of whales, he concludes that it is not committed to mammals. But when an essentialist applies it, the conclusion is the opposite.

Formulations such as (1) and (2) reveal the influence (and importance) of a period that featured intense work in the areas of extensional formal semantics and the axiomatization of both scientific and philosophical theories. Such criteria are fairly characterized as seeing existential quantification as the source (and signal) of ontological commitment. They may be seen as differing elaborations of the theme expressed in Quine's famous slogan, 'To be is to be a value of a variable' (1939). Quine would later adopt a somewhat different perspective on the matter, the perspective of 'ontological relativity', which we will mention below. But there remain two very general criticisms of efforts in the quantificational spirit that merit separate discussion.

3 Two further problems

For a time the influence of Tarskian formal semantics was so great that it must have been difficult to imagine that any essentially different treatment of quantification could be defensible. But there is a coherent alternative called the substitutional interpretation. On this interpretation the semantics of a first-order language includes a non-empty set of terms of that language called the 'substitution class'. An existential quantification is counted true if and only if there is a member of the substitution class such that the result of removing the quantifier and replacing all resulting free occurrences of its variable by that member is true. So the truth-conditions for the quantifiers make no appeal to a domain of (potentially) extra-linguistic entities. Depending on how truth-values are assigned to atomic sentences - and one possibility is by sheer stipulation - a given substitutionally interpreted theory may have no intuitive ontological commitment beyond the constituents of the language itself, the truth-values, and perhaps some simple sets (including functions) of these entities.

The usefulness of substitutional quantification for representing informal theories may be debated (see [Quantifiers, substitutional and objectual](#)). One may ask whether the concept of truth embodied in a given substitutional theory appropriately matches the one implicit in a given informal theory. If the intuitive verdict is that the informal theory has non-trivial extra-linguistic commitment, then a substitutional theory lacking such commitment will be metaphysically inadequate (though it might get the truth-values right). But there remains the simple fact that the substitutional interpretation is perfectly coherent. From this it appears to follow that the mere assertion of an existential sentence does not provide a sufficient condition for non-trivial ontological commitment. So an adequate criterion of commitment should not presuppose that it does. The very existence and coherence of the substitutional interpretation thus creates a real difficulty for 'quantificational' criteria in the spirit of (1) or (2).

It has often been claimed, for one reason or another, that commitment is an 'intensional' notion (see [Intensional entities §1](#)). The main reason why (2) improves upon (1) is that it contains the word 'must'. But this suggests that the improvement is gained only at what Quine must see as great cost, namely, by exploiting the notion of *necessity*, which belongs to the theory of meaning. For it is natural to interpret the condition stated in (2) as meaning that it is necessary that if the statements affirmed in the theory are true, then objects of the required sort are values of its variables.

Others have found intensionality not in an appeal to necessity but in the (claimed) 'referential opacity' of the commitment relation itself, that is, in the resistance of its second position both to the substitution of coreferential expressions and to existential generalization. Commitment to Twain does not entail commitment to Clemens; nor does commitment to unicorns (or horses) entail there being entities to which one is committed. (Note that a

plausible claim of opacity can only be made concerning *prima facie* commitment.)

Here is a further (but related) reason to suspect intensionality. Suppose we try to state a criterion in which the condition expresses a binary relation between persons or theories (and so on) and something else. What are the plausible candidates for the second relatum? Consider first the theory that simply says there are horses. Well, there *are* horses, but the theory is not committed to any of them (either individually or collectively). One natural thought is that the second relatum is the property (or kind), being a horse. Accordingly, to say that the theory is committed to horses would mean that it cannot be true unless the property is instantiated. But properties are intensional entities. (In the case of commitment to specific entities, ‘possible objects’, including *merely* possible objects, are natural candidates analogous to properties.) There are other possibilities, but it has been argued that any adequate relational treatment must be intensional.

Yet another approach sees theories not as linguistic or quasi-linguistic entities, whether formal or informal, but rather as propositions (or sets of propositions). Then a theory is committed to, say, unicorns if it entails the proposition that unicorns exist. This approach is intensional because propositions are intensional entities. (It also appeals to entailment, which is usually (though not universally) thought to rest on necessity.)

So the suspicion of intensionality may arise for a number of reasons. But the intensionality always seems to flow from one of two somewhat related sources. The first is the classical problem of explaining how we can either assert or deny the existence of things that in fact do not exist. The other is that we are treating *prima facie* rather than all-in commitment. The (claimed) intensionality of commitment is not in itself disturbing, but it was vexing for Quine because of his general aversion to the intensional.

4 Translation and triviality

The possibility that an adequate criterion of commitment would somehow be ‘trivial’ arose in §1. Remarkably enough, it also surfaced in Quine in the wake of his important doctrines of ‘ontological relativity’ and the ‘inscrutability of reference’. Details of these positions cannot be discussed here, but their effect on the matter of a criterion should be mentioned.

The key effect is to cast the question of a criterion as one concerning *translation*. Theories are stated in languages. An ontological criterion must also be stated in a language. For our purposes, the language of a criterion may be viewed as a technically enhanced version of an appropriate fragment of English. Call such a language E^* . Now it seems clear that if the question whether a theory T is committed to entities of kind K can be settled at all, then it can be settled as follows: simply translate the assertions (or theorems) of T into E^* and see whether the results include ‘There exist objects of kind K ’ (where ‘There exist’ is understood referentially).

For reasons flowing from the aforementioned doctrines, there is a sense in which Quine does not regard this as ‘settling’ the question of T ’s ontology, because he thinks ‘There exist objects of kind K ’, in E^* , is itself ontologically indeterminate. But he also thinks that this is an indeterminacy that cannot be removed and should not be lamented.

Any such ‘translational’ criterion can easily seem trivial. For either we already know the commitments of T , and so have no need for it, or else we do not, and so cannot apply it. Yet it seems perfectly natural and correct independently of the Quinean doctrines and, in particular, independently of their implication that it does not settle the question of a theory’s commitment.

See also: [Ontology](#); [Free logics, philosophical issues in](#); [Logical and mathematical terms, glossary of](#)

MICHAEL JUBIEN

References and further reading

The following items all involve complex argumentation and at least some technicality. Many contain valuable further references.

Cartwright, R.L. (1954) ‘Ontology and the Theory of Meaning’, *Philosophy of Science* 21: 316-25. (The classic statement of the idea that ontological commitment involves the notion of necessity, as discussed in §3.)

Dunn, J.M. and Belnap, N.D. (1968) ‘The Substitution Interpretation of the Quantifiers’, *Noûs* 2: 179-85. (A discussion and technical presentation of substitutional semantics, as discussed in §3.)

- Jubien, M.** (1972) 'The Intensionality of Ontological Commitment', *Noûs* 6: 378-87. (Argues that any adequate criterion of commitment that treats the notion as a relation must be intensional, as mentioned in §3.)
- Kripke, S.A.** (1976) 'Is there a Problem about Substitutional Quantification?', in G. Evans and J. McDowell (eds) *Truth and Meaning*, Oxford: Clarendon Press, 325-419. (Argues that attempts to see substitutional quantification as covertly carrying the commitments of referential quantification are incorrect. Very detailed and technical in places. See the present §3.)
- Marcus, R.B.** (1972) 'Quantification and Ontology', *Noûs* 6: 240-50; repr. in *Modalities*, Oxford and New York: Oxford University Press, 1993, 75-87. (A classic statement of the ontological neutrality of substitutional quantification, as discussed in §3.)
- Marcus, R.B.** (1978) 'Nominalism and the Substitutional Quantifier', *Monist* 61: 351-62; repr. in *Modalities*, Oxford and New York: Oxford University Press, 1993, 111-24. (Explores the congeniality of substitutional quantification for various forms of nominalism.)
- Quine, W.V.** (1939) 'A Logistical Approach to the Ontological Problem', presented at the fifth International Congress for the Unity of Science, Cambridge, MA; repr. in *The Ways of Paradox and Other Essays*, New York: Random House, 1966, 64-9. (An early discussion of naming, reference and ontology. The source of 'To be is to be a value of a variable' mentioned in §2.)
- Quine, W.V.** (1953a) 'Notes on the Theory of Reference', in *From a Logical Point of View*, Cambridge, MA and London: Harvard University Press, 1961, 130-8. (The source of criterion (1), discussed in §§1-3, and of an explicit claim that commitment belongs to the theory of reference.)
- Quine, W.V.** (1953b) 'Logic and the Reification of Universals', in *From a Logical Point of View*, Cambridge, MA and London: Harvard University Press, 1961, 102-29. (The source of criterion (2), discussed in the present §§1-3. Contains an explicit claim of the need for a criterion of commitment.)
- Quine, W.V.** (1968) 'Ontological Relativity', *Journal of Philosophy* 65: 185-212; repr. in *Ontological Relativity and Other Essays*, New York and London: Columbia University Press, 1969, 26-68. (An early statement of the doctrine of ontological relativity and related matters, as mentioned in §4.)
- Quine, W.V.** (1981) 'Things and Their Place in Theories', in *Theories and Things*, Cambridge, MA and London: Belknap Press, 1-23. (More on ontological relativity, including the idea that the question of the ontology of a theory is fundamentally a question concerning translation. See the present §4.)
- Scheffler, I. and Chomsky, N.** (1958-9) 'What is Said to Be', *Proceedings of the Aristotelian Society* 69: 215-24. (The classic statement of the idea that the commitment context is referentially opaque; discussed in §3.)
- Schwartz, R.** (1993) 'On "What is Said to Be"', *Synthese* 94: 43-54. (A contemporary attempt to capture a concept of commitment using only extensional resources. See the present §§3, 4.)

Ordinal logics

By an ordinal logic is meant any uniform effective means of associating a logic (that is, an effectively generated formal system) with each effective ordinal representation. This notion was first introduced and studied by Alan Turing in 1939 as a means to overcome the incompleteness of sufficiently strong consistent formal systems, established by Kurt Gödel in 1931.

The first ordinal logic to consider, in view of Gödel's results, would be that obtained by iterating into the constructive transfinite the process of adjoining to each system the formal statement expressing its consistency. For that ordinal logic, Turing obtained a completeness result for the class of true statements of the form that all natural numbers have a given effectively decidable property. However, he also showed that any ordinal logic (such as this) which is strictly increasing with increasing ordinal representation cannot have the property of invariance: in general, different representations of the same ordinal will have different sets of theorems attached to them. This makes the choice of representation a crucial one, and without a clear rationale as to how that is to be made, the notion of ordinal logic becomes problematic for its intended use.

Research on ordinal logics lapsed until the late 1950s, when it was taken up again for more systematic development. Besides leading to improvements of Turing's results in various respects (both positive and negative), the newer research turned to restrictions of ordinal logics by an autonomy (or 'boot-strap') condition which limits the choice of ordinal representations admitted, by requiring their recognition as such in advance.

1 Incompleteness, consistency statements and reflection principles

The point of departure for the study of ordinal logics is Kurt Gödel's second incompleteness theorem, also referred to as the theorem on unprovability of consistency (see [Gödel's theorems](#)). With each effectively generated formal system S containing a modicum of arithmetic is associated the arithmetical statement Con_S of consistency of S . If S is consistent then Con_S is not provable in S , and so the system S' obtained by adjoining Con_S as an axiom to S is stronger than S . It may, in fact, happen that S' is inconsistent, but if S satisfies a condition slightly stronger than consistency (to be described below) so also does S' . Let $\mathcal{C}$ be the class of effectively generated formal systems satisfying this strengthened consistency condition. Thus by beginning with any system S_0 in $\mathcal{C}$ and forming, in succession, $S_1 = S'_0, S_2 = S'_1, \dots$, we obtain more and more complete systems in $\mathcal{C}$. As will be explained in §2, this procedure can also be extended into the transfinite, since the union of any effectively generated sequence of systems in $\mathcal{C}$ is again in $\mathcal{C}$. In this section, matters concerning formal systems and extension principles such as those of Gödelian type will be explained as needed to make these notions precise. Certain preliminaries from recursion theory and arithmetical definability theory are necessary for this purpose (see [Computability theory](#); [Recursion-theoretic hierarchies](#)).

The variables $x, y, z, \dots$ are here taken to range over the set $N = \{0, 1, 2, \dots\}$ of natural numbers. For $P(x)$ a property of natural numbers, $\{x: P(x)\}$ denotes the set of all x in N for which $P(x)$ holds. The class Σ_1^0 in the arithmetic hierarchy consists of all sets of the form $\{x: \exists y R(x, y)\}$ where R is primitive recursive; Σ_1^0 is coextensive with the class of recursively enumerable (RE) sets, i.e. those which are effectively generated according to Church's thesis. There is a uniform Σ_1^0 enumeration $W_e = \{x: yT(e, x, y)\}$ of all Σ_1^0 sets for $e = 0, 1, 2, \dots$, where T is a suitable primitive recursive relation; e is said to be an *index* of an effectively generated set W if $W = W_e$. The only other classes in the arithmetic hierarchy to which reference will be made below are the Π_1^0 sets $\{x: \forall y R(x, y)\}$ and the Π_2^0 sets $\{x: \forall y \exists z R(x, y, z)\}$. (See [Recursion-theoretic hierarchies](#) for more information about the Π and Σ classifications.)

For simplicity, on the formal side we shall deal only with systems whose language is that of first order (or elementary) number theory. Primitive recursive relations may be taken as basic, or defined (by a method due to Gödel) in terms of the operations of addition and multiplication. Either way, we can then speak of formulas in Σ_1^0 , Π_1^0 and Π_2^0 form (among others). All notions of truth and definability are referred to the standard model, in which the variables are taken to range over N and the basic operations and relations receive their intended interpretation.

To carry notions of effectiveness over to the syntactic side, one uses the association with each formula A of its

Gödel number, denoted $\#A$. Then the above enumeration of $\sum_1^0$ sets induces an enumeration of sets of formulas, $S[e] = \{A: \#A \in W_e\}$; a set S of formulas is said to be effectively generated (or RE) if $S = S[e]$ for some e , and in that case e is called an *index* or *presentation* of S . Effective operations on such sets are explained in terms of operations on their indices. Moreover, effective sequences $S_n (n \in \mathbb{N})$ of RE sets of formulas can be considered as given by (an index f of) a total recursive function $\{f\}$ with $S_n = S[\{f\}(n)]$ for each n . Associated effectively with each such f is an index of $\cup_n S_n$ as an RE set.

Gödel's incompleteness theorems apply to effectively generated extensions of a formal system containing a modicum of arithmetic. Most familiar for the latter purpose is the system PA of Peano Arithmetic (Peano's axioms in first-order form), though much weaker systems suffice. Thus in the following, only RE sets S of formulas will be considered for which S includes PA . Then A is provable from S if A is a consequence of S by means of the axioms and rules of the first-order predicate calculus with identity. For simplicity, we assume that the non-logical axioms of S are given as closed formulas, that is, sentences. S is *consistent* if it does not prove the sentence $0 \neq 0$, and is *1-consistent* if each $\sum_1^0$ sentence provable in S is true. S is called *ω -consistent* if there is no formula $A(x)$ for which we have both $\exists x A(x)$ and $\neg A(\bar{n})$ provable in S for each n . (The notion of ω -consistency was introduced by Gödel in 1931 to show that if G is the sentence which expresses of itself that it is not provable in S then $\neg G$ is not provable in S ; it was observed later that 1-consistency suffices for this conclusion.) Finally, S is said to be *true* if each sentence of S is true. Note that truth implies ω -consistency which implies 1-consistency which, in turn, implies consistency. We write $S_1 S_2$ if every A provable in S_1 is provable in S_2 , and $S_1 \equiv S_2$ if $S_1 \subseteq S_2$ and $S_2 \subseteq S_1$. S_2 is said to be *stronger* than S_1 if $S_1 \subseteq S_2$ but some A is provable in S_2 which is not provable in S_1 .

With each index or presentation e of an effectively generated system S is associated a $\sum_1^0$ formula $Prov_S(x)$ which expresses that x is the number of a formula provable from S . Then the $\prod_1^0$ sentence $Con_S = \neg Prov_S(\#A_0)$, where A_0 is $(0 \neq 0)$, formally expresses the consistency of S . (To be more precise, we should write $Prov_e(x)$ and Con_e for these formulas, for each index e of S .) According to Gödel's second incompleteness theorem, if S is consistent (and includes PA) then S does not prove Con_S . By hypothesis, Con_S is true in this case, but it may be that $S' = S \cup \{Con_S\}$ is inconsistent; however, if S is 1-consistent then so also is S' .

For familiar systems S , we come to accept their consistency by informally recognizing that each axiom of S is true, hence that each sentence A provable from S is true. The formal statement of the latter recognition can be given in the language of arithmetic without the truth predicate, simply as

$$(*) \text{ } Prov_S(\#A) \rightarrow A$$

For, by Tarski's truth criterion (see **Tarski's definition of truth**), we have $Tr(\#A) \leftrightarrow A$. The collection of all statements of the form $(*)$ is called the *(local) reflection principle* (Rfn_S) for S . Note that the instance $Prov_S(\#A_0) \rightarrow A_0$, with A_0 as above, is equivalent to $\neg Prov_S(\#A_0)$, that is, to Con_S . In other words, the consistency of S is a consequence of Rfn_S . As with the addition of Con_S to S , 1-consistency is preserved under the expansion of S by all instances of Rfn_S for which A is in $\prod_2^0$ form. However, we must assume a stronger hypothesis for the full extension, namely: if S is true so also is $S \cup Rfn_S$. In the following, we shall be concerned with the extensions of systems given by $S' = S \cup \{Con_S\}$ and $S' = S \cup Rfn_S$, among others. In each of these cases, an index of S' is found effectively from one for S .

2 Ordinal logics; Turing's results

Passages from S to S' for effectively generated S , such as described in the preceding section, can be iterated effectively into the transfinite by using closure under effective unions at limit stages. For this one needs a notion of a system of recursive (or 'constructive') ordinal notations. The first such system was introduced in 1936 by Alonzo Church and Stephen C. Kleene using representations of ordinals by certain terms in Church's lambda calculus (see **Lambda calculus**). Several variants of this system were subsequently introduced which are all equivalent in the sense that the same ordinals are represented by the different systems. One of these was given by Kleene (1938) in more current recursion-theoretic terms as a subset O of $\mathbb{N}$ together with a partial ordering relation $<_O$ between elements of O and a map $a \mapsto |a|$ from O into the ordinals, with the property that if $a <_O b$ then $|a| < |b|$.

This representation is not in general unique, that is, the same ordinal α may have many $a \in O$ with $|a| = \alpha$ (if there are any such $a \in O$ at all). The relation $<_O$ is tree-like in the sense that for each $b \in O$, $\{a: a <_O b\}$ is linearly ordered by $<_O$.

The system O is also equipped with: (i) a unique element 0_O with $|0_O| = 0$, (ii) a recursive operation sc_O from O to O such that for $a' = sc_O(a)$, $|a'| = |a| + 1$ and (iii) a recursive operation lim_O from indices of recursive increasing sequences of members of O to O , such that whenever $\{f\}(n) = a_n$ is in O and $a_n <_O a_{n+1}$ for all n then $|lim_O(f)| = lim_n |a_n|$. Finally, O is the least set satisfying the recursive closure conditions (i) - (iii). An ordinal α is said to be recursive if $\alpha = |a|$ for some $a \in O$; in a sense, these are the effectively countable ordinals. The least non-recursive ordinal is denoted ω_1^{CK} , i.e. the least effectively uncountable ordinal (in the sense of Church and Kleene).

Turing's notion of ordinal logic in his seminal 1939 paper was framed in terms of the lambda calculus; moreover, his notion of a logic was idiosyncratic in that he only considered sets of Π_2^0 formulas. Basically, for Turing, an ordinal logic is any effective association of a logic in his sense with each constructive ordinal notation. The following is an adaptation of Turing's notions to the more current recursion-theoretic framework together with the treatment of logics as formal systems in the sense of §1, which respects all of his results.

Thus, by an *ordinal logic* is meant any partial recursive E which associates with each $a \in O$ an index $E(a)$ of an effectively generated formal system S_a , so that $S_a = S[E(a)]$. In the rest of this section S_a is assumed to be a set of sentences in the language of arithmetic with $PA \subseteq S_a$. The logic given by E is denoted by $\langle S_a^E \rangle_{a \in O}$, and the superscript ' E ' is suppressed when there is no ambiguity. An ordinal logic is said to be *invariant (up through α)* if for all $a, b \in O$ with $|a| = |b|$ and $|b| \leq \alpha$ we have $S_a \equiv S_b$; it is said to be *increasing (strictly increasing, up through α)* if whenever $a <_O b$ then $S_a \subseteq S_b$ (S_b is stronger than S_a , for $|b| \leq \alpha$). The ordinal logic is said to be *continuous* if whenever f represents an increasing sequence in O , then $S_{lim_O(f)} = \cup_n S_{\{f\}(n)}$. Finally, it is said to be *complete* for a class T of sentences if for each A in T there exists $a \in O$ with A provable in S_a .

Of special interest are ordinal logics obtained in the following way. We suppose given a partial recursive function (called a *succession principle*) which associates with each index of an effectively generated system S an index of a system S' extending S . Then given any initially specified system S_0 containing PA , one can construct an ordinal logic $\langle S_a \rangle_{a \in O}$ such that:

(i) $S_{0_O} = S_0$, (ii) $S_{sc_O(a)} = S'_a$ for each $a \in O$, and (iii) the ordinal logic is continuous. Then given any notion of correctness $\mathcal{C}$ of systems which applies to S_0 and which is closed under the operation $S \mapsto S'$ and under effective unions, each S_a will belong to $\mathcal{C}$. This holds for the collection $\mathcal{C}$ of 1-consistent systems when $S' = S \cup \{Con_S\}$ and for the collection of true systems when $S' = S \cup Rfn_S$. An ordinal logic constructed satisfying (i)-(iii) is said to be based on the succession principle $S \mapsto S'$; it is strictly increasing if S' is always stronger than S , for S in $\mathcal{C}$.

For the ordinal logic based on the succession principle $S' = S \cup Rfn_S$, Turing obtained completeness for the class T of true Π_1^0 sentences. He asked whether one can obtain completeness for the class of true Π_2^0 sentences; this was answered in the negative by **Feferman** in 1962 (see §3). In any case, Turing's Π_1^0 completeness result is disappointing (as he himself thought) due to the following general result which is obtained: an ordinal logic cannot be strictly increasing and invariant. Thus a true Π_1^0 sentence provable in some S_a need not be provable in an S_b with $|b| = |a|$, and so the completeness result is sensitive to which ordinal notations are used.

Closer inspection of Turing's arguments revealed the following. First of all, his completeness result for true Π_1^0 sentences already holds for the ordinal logic based on the succession principle $S' = S \cup \{Con_S\}$. For this, one associates with each primitive recursive predicate $R(x)$ a number a_R such that a_R is in O just in case $\forall x R(x)$ is true, and is such that $|a_R| = \omega + 1$ and S_{a_R} proves $\forall x R(x)$ when that holds. The notation a_R is not 'natural', unlike our usual notation for $\omega + 1$; it is simply fabricated to do this job. Similarly, an ordinal logic can't be strictly increasing and invariant up through $\omega + 1$; this is best possible for continuous ordinal logics, since one has unique notations in O for the natural numbers, hence invariance up through ω .

The original aim for ordinal logics was to obtain an association of effective formal systems S_α with ordinals α in

such a way that $S_{\alpha+1} = S'_\alpha$ and $S_\alpha = \cup_{\beta < \alpha} S_\beta$ for limit α . The definition of ordinal logics as assignments $a \mapsto S_a^E$ for $a \in O$ was necessitated in order to make the association effective. Unless one has some canonical way of associating with each recursive ordinal α an $a \in O$ with $|a| = \alpha$, the notion of ordinal logic is problematic for its original purpose.

3 Subsequent work

Turing did no further work on ordinal logics after 1939, and the subject went into limbo until it was taken up again in the late 1950s by Georg Kreisel and by Feferman. Part of Feferman's work (1962) consisted in recasting Turing's work from the lambda calculus in ordinary recursion-theoretic terms and in sharpening his results, essentially as described in the preceding section. Finally, attention was restricted to continuous ordinal logics based on a given effective succession principle $S \mapsto S'$; these were called *transfinite recursive progressions of axiomatic theories*. The new work in 1962 was for a recursive progression based on the so-called *uniform reflection principle for a logic S* (RFN_S) which consists of all sentences of the form

$$(**) \quad \forall x \text{Prov}_S(\#A(\bar{x})) \rightarrow \forall x A(x)$$

where for each n , $\bar{n}$ is the numeral denoting n . Thus (RFN_S) for $A(x)$ expresses that if S proves $A(\bar{n})$ for each n then $\forall x A(x)$ is true, and so it is a kind of formalized ω -rule. The main result of the author's 1962 work was that for the recursive progression based on the uniform reflection principle, we have completeness for *all* true arithmetical sentences. Moreover, one obtains completeness along suitable paths P through O (i.e. where P is a subset of O linearly ordered by $<_O$ which contains a representative for each recursive ordinal). Namely, it was shown that there is a path P through O (and recursive in O) such that for each true arithmetical sentence A , there exists $a \in P$ with A provable in S_a . Now this completeness result is sensitive to the choice of path because of Turing's non-invariance result, and so the crucial question to be raised is: *what is a natural path P through O?* As of the time of writing, there is still no convincing answer to this question. Familiar natural systems of ordinal representation all close off at a recursive ordinal. One feature of these is that they are defined inductively; furthermore, sets generated by arithmetical inductive definitions are in the class $\prod_1^1$ in the analytic hierarchy (as is the set O). (See [Recursion-theoretic hierarchies](#) for more information about the analytic hierarchy.) It was shown by Feferman and Spector in 1962 that $\prod_1^1$ paths through O exist, but that if P is any such path then $\cup_{a \in P} S_a$ is incomplete for true $\prod_1^0$ sentences. Hence if the answer to the question about natural paths requires them to be inductively defined, one will have a very strong incompleteness result for recursive progressions restricted to natural ordinal notations.

4 Autonomous progressions of theories and reflective concepts of proof

In order to characterize informal concepts of proof embodying some kind of stepwise reflective character such as that of *finitist proof* or *predicative proof*, Georg Kreisel suggested in 1958 the use of certain ordinal logics restricted by an *autonomy* (or 'boot-strap') condition. The restriction consists in allowing one to advance to an S_a for $a \in O$ only if it has been proved in some previously accepted S_b ($b <_O a$) that a is in O . Formally, this requires an extension of the language of arithmetic to admit predicates expressing well-foundedness in one way or another. Not so incidentally, imposition of the autonomy condition partially circumvents the problem of which ordinal notations are to be used in an ordinal logic. However, the purposes are substantially different, since autonomous progressions are incomplete for true $\prod_1^0$ sentences and break off at a recursive ordinal.

For the concept of predicative provability in analysis, Kreisel proposed consideration of an autonomous progression of ramified analytic systems. In 1964, Kurt Schütte and Feferman independently obtained a characterization of the least non-predicatively provable ordinal under this proposal. In a 1991 article, Feferman introduced a notion of *reflective closure* of a system S which does not require any of the (problematic) machinery of ordinal logics and showed that the reflective closure of PA has exactly the same arithmetical consequences as the autonomous progression of ramified systems. It is proposed there that the reflective closure of a system S contains everything one ought to accept if one has accepted the basic notions and principles of S .

See also: [Logical and mathematical terms, glossary of](#)

SOLOMON FEFERMAN

References and further reading

- Davis, M.** (ed.) (1965) *The Undecidable: Basic Papers on Undecidable Propositions, Unsolvable Problems and Computable Functions*, Hewlett, NY: Raven Press. (Includes a reprinting of Turing's 1939 paper on ordinal logics, among other fundamental papers by Gödel, Church, Kleene, Turing and Post.)
- Feferman, S.** (1962) 'Transfinite Recursive Progressions of Axiomatic Theories', *Journal of Symbolic Logic* 27: 259-316. (Re-examination, technical improvement and extension of Turing's work on ordinal logics as described in §3.)
- Feferman, S.** (1964) 'Systems of Predicative Analysis', *Journal of Symbolic Logic* 29: 1-30. (Use of autonomous progressions of theories for the characterization of predicate provability in analysis, referred to in §4.)
- Feferman, S.** (1988) 'Turing in the Land of $O(z)$ ', in R. Herken (ed.) *The Universal Turing Machine: A Half-Century Survey*, Oxford: Oxford University Press, 113-47. (A readable account of how Turing came to work on ordinal logics, together with a semi-technical exposition of his basic results and of subsequent work.)
- Feferman, S.** (1991) 'Reflecting on Incompleteness', *Journal of Symbolic Logic* 56: 1-49. (Introduces the concept of reflective closure and shows its equivalence with the autonomous progression for predicative provability; referred to in §4.)
- Feferman, S. and Spector, C.** (1962) 'Incompleteness Along Paths in Progressions of Theories', *Journal of Symbolic Logic* 27: 383-90. (The incompleteness results described in §3.)
- Gödel, K.** (1931) 'Über formal unentscheidbare Sätze der *Principia Mathematica* und verwandter Systeme I', *Monatshefte für Mathematik und Physik* 38: 173-98; trans. 'On Formally Undecidable Propositions of *Principia Mathematica* and Related Systems', in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967, 592-617. (Gödel's fundamental paper on incompleteness of sufficiently strong formal systems; reprinted in the original and in English translation in [Gödel 1986](#).)
- Gödel, K.** (1986) *Collected Works*, vol. 1, *Publications 1929-1936*, ed. S. Feferman, J.W. Dawson Jr, S.C. Kleene, G.H. Moore, R.M. Solovay, and J. van Heijenoort, New York and Oxford: Oxford University Press.
- Hodges, A.** (1993) *Alan Turing: The Enigma*, London: Burnett Books; New York: Simon & Schuster. (A very readable, absorbing and informative biography of Turing with a description of all his work.)
- Kleene, S.C.** (1938) 'On Notation for Ordinal Numbers', *Journal of Symbolic Logic* 3: 150-5. (Kleene's system O of recursive ordinal notations described in §2.)
- Kreisel, G.** (1960) 'Ordinal Logics and the Characterization of Informal Concepts of Proof', *Proceedings of the International Congress of Mathematicians, 14-21 August 1958*, Cambridge: Cambridge University Press, 289-99. (Proposed the use of autonomous ordinal logics for the characterization of finitist proof and suggested a similar use for predicative proof; referred to in §4.)
- Kreisel, G.** (1970) 'Principles of Proof and Ordinals Implicit in Given Concepts', in J. Myhill, R.E. Vesley and A. Kino (eds) *Intuitionism and Proof Theory*, Amsterdam: North Holland, 489-516. (Discussion of the work on autonomous ordinal logics for finitist and predicative provability.)
- Turing, A.M.** (1939) 'Systems of Logic Based on Ordinals', *Proceedings of the London Mathematical Society*, series 2, 45: 161-228; repr. in M. Davis (ed.) *The Undecidable*, Hewlett, NY: Raven Press, 1965, 155-222. (Turing's fundamental paper on ordinal logics whose notions and results are described in §2.)

Paraconsistent logic

A logic is paraconsistent if it does not validate the principle that from a pair of contradictory sentences, A and $\sim A$, everything follows, as most orthodox logics do. If a theory has a paraconsistent underlying logic, it may be inconsistent without being trivial (that is, entailing everything). Sustained work in formal paraconsistent logics started in the early 1960s. A major motivating thought was that there are important naturally occurring inconsistent but non-trivial theories. Some logicians have gone further and claimed that some of these theories may be true. By the mid-1970s, details of the semantics and proof-theories of many paraconsistent logics were well understood. More recent research has focused on the applications of these logics and on their philosophical underpinnings and implications.

The idea that a contradiction implies everything (*ex contradictione quodlibet* - ECQ) has always been a contentious one in logic. Despite this, formal paraconsistent logics, which do not validate ECQ, are creatures of the twentieth century. The earliest ones were constructed in Russia by N.A. Vasil'ev c.1912 - an Aristotelian logic - and I.E. Orlov in 1929 - a relevance logic (see [Anderson, Belnap and Dunn 1992](#)). However, these had no impact at the time.

Work on formal paraconsistent logics did not begin in earnest until after the Second World War. Since then they have been proposed independently by many logicians, the earliest notable ones being S. Jaśkowski (in Poland) in 1948, F.G. Asenjo (Argentina) c.1954, N.C.A. da Costa (Brazil) c.1958 and T.J. Smiley (the UK) in 1959. Work on relevance logic by [A.R. Anderson and N.D. Belnap](#) (in the USA) also started in the late 1950s (see [Relevance logic and entailment](#)), and the specifically paraconsistent aspects of relevance logic were developed by R. Routley and others (in Australia) in the late 1960s and 1970s. (Note that a paraconsistent logic need not be relevant.) Since then, work on formal paraconsistent logic has continued apace in many places, but most notably in Brazil (under the leadership of da Costa) and Australia. (Some of the original work is rather difficult to come by or of a rather preliminary nature. The best access points in the literature are: Jaśkowski (1969); Asenjo (1966); da Costa (1974); Anderson and Belnap (1975); and Routley (1977).)

A major motivation behind the construction of formal paraconsistent logics has always been the idea that in many contexts we may have information that is inconsistent, but from which we want to draw conclusions in a controlled way. (The term 'paraconsistent' was coined by M. Quesada at the third Latin American Symposium on Mathematical Logic in 1976, to indicate just this.) Examples that are frequently appealed to are: evidence provided by different witnesses, constitutions and other legal documents, various scientific theories, numerous philosophical theories, and information in a computer database. In such contexts, even though the data are incorrect, if we are stuck with them (as we may well be), then the logic had better be paraconsistent. Moreover, inconsistent scientific theories, even if they are not correct, may still be useful, or good approximations to the truth.

Some paraconsistent logicians have claimed that inconsistent theories may actually be true (see, for example, [Priest 1987](#)). The view that some contradictions or contradictory theories are true is called 'dialeth(e)ism' - from 'dialetheia' (a term coined by Priest and Routley in 1982), which means a true statement of the form $A \ \& \ \sim A$. The most commonly cited examples of dialetheias are the paradoxes of self-reference, such as Russell's and the liar paradox (see [Paradoxes of set and property](#); [Semantic paradoxes and theories of truth](#)). The failure to obtain consensus on any consistent account of the paradoxes gives this suggestion its appeal. Other suggested examples of dialetheias include: statements about objects on the borderline of some vague predicate; moral dilemmas; and dialectical contradictions in the tradition of Hegel and Marx.

Some approaches to paraconsistent logic, such as Smiley's, obtain a suitable inference relation by starting with the classical one and filtering out ECQ and other undesirables. Such logics typically give up the transitivity of entailment. A more common approach is to specify a notion of entailment semantically, defined in terms of truth-preservation over a class of interpretations. For this approach, it is necessary to have a mechanism whereby contradictory sentences may simultaneously hold in an interpretation. For this reason, one may think of paraconsistent logic as a kind of dual of intuitionist logic (see [Intuitionism](#)), the former violating the law of non-contradiction, the latter violating the law of the excluded middle. (Though it is quite possible to have a paraconsistent logic in which $\sim (A \ \& \ \sim A)$ is semantically valid.) Paraconsistent logics of this kind

characteristically invalidate the Disjunctive Syllogism (DS): $A \vee B$ and $\sim A$ entail B . For both premises may be true in virtue of the properties of A , while B is not.

Various techniques have been proposed to achieve the required end. Jaśkowski's is to interpret 'true' as 'true in some possible world or other'. Da Costa's is to give up the truth-functionality of negation, so that if A is true, $\sim A$ may be either true or false. Routley's is to treat negation as an intensional operator, so that $\sim A$ is true at a world w if A is false at some associated world, w^* . Asenjo's suggestion (which can also be harnessed in the semantics of relevance logics) is to allow sentences to take a non-classical truth value, which may be thought of as *both true and false*, and which is a fixed point for negation.

By the mid-1970s the semantics and proof theories of many paraconsistent logics were well developed. More recently, much work has gone into their applications, both technical and philosophical. The technical applications all involve the investigation of inconsistent theories. Notable results in this area include a proof that naïve set theory with (an unrestricted comprehension axiom and) a suitable underlying paraconsistent logic is non-trivial (Brady 1989), and a proof (by R. Meyer) that there is a (consistent!) arithmetic that can prove its own non-triviality. The techniques involved in the latter involve the construction of models of full first-order arithmetic, many of which are finite (see Meyer and Mortensen 1984). It is also possible to construct non-trivial theories that contain both self-reference and epistemic operators, and that are semantically closed, in the sense of Tarski (see Priest 1991a).

The philosophical applications are more diffuse. Traditionally, consistency was thought to be the cornerstone of many important philosophical notions (for example, truth, rationality), but the viability of paraconsistent logic throws down a challenge to any such claim. In the light of this, a reappraisal both of such notions and of the significance of results that turn on consistency, for example, Gödel's incompleteness theorems, is called for and is in train.

Many would be prepared to concede the possibility of a limited use for paraconsistent logic, for example, as an inference engine for a computational database. The major criticism as far as this goes is that a paraconsistent logic is too weak to permit useful inference. In particular, a number of writers have argued that the DS is essential to most practical inference. In the author's opinion, this sort of objection carries little weight. For a start, the logical resources of the programming language PROLOG are validated in most paraconsistent logics when suitably interpreted. More generally, most paraconsistent logicians have been prepared to accept the usability of principles such as the DS in consistent ('normal?') contexts. This idea has motivated the construction of formal non-monotonic logics in which the DS is a default inference (see Batens 1989; Priest 1991b).

Much of the criticism of paraconsistent logic has fallen on dialetheism. One of the chief general criticisms is that the paraconsistent semantics for negation fails to capture (true) negation, but this is difficult to make stick for the many-valued semantics. Some have argued that if contradictions are not to be rejected as logically unacceptable then logic is ruined as an instrument of rational criticism, but this objection trades on a confusion of what is possible according to formal logic and what is rationally possible. A third important general criticism is that if one is to accept A one must rationally reject $\sim A$. However, dialetheism aside, there are situations where we seem to have little rational option but to accept a contradiction (for example, the paradox of the preface). Most of the application-specific criticisms have tried to undercut a dialethic solution to the self-referential paradoxes by arguing that paraconsistent solutions to these collapse into triviality - or, at least, fail to do so only by *ad hoc* manoeuvres of a kind that make consistent purported solutions so unsatisfactory. To date, such arguments have not hit the mark squarely. (On much of the above, see Smiley and Priest (1993).)

The possibility of violations of the law of non-contradiction was widely canvassed by pre-Aristotelian philosophers. Aristotle attacked this possibility in *Metaphysics* (1005b8-1009a5). His arguments are distinctly dubious, as was shown by Łukasiewicz (1971), but his authority has determined subsequent orthodoxy. Only a few (notably Hegel) have challenged it. Even fewer have produced a sustained defence of the law. The technical viability of paraconsistent logic and dialetheism therefore raises a challenge to contemporary philosophy of no little significance.

See also: [Logical and mathematical terms, glossary of](#)

GRAHAM PRIEST

References and further reading

Many of the works cited in §1 are in rather inaccessible places, not in English or of a preliminary nature. I have often therefore cited, instead, later papers which do not have the same features. All the works cited contain both formal and non-formal material, the relative amounts varying.

- Anderson, A.R. and Belnap, N.D.** (1975) *Entailment*, vol. 1, Princeton, NJ: Princeton University Press. (A summary of the early work on relevance logic.)
- Anderson, A.R., Belnap, N.D. and Dunn, J.M.** (1992) *Entailment*, vol. 2, Princeton, NJ: Princeton University Press. (Includes references to Vasil'ev and Orlov.)
- Asenjo, F.G.** (1966) 'A Calculus of Antinomies', *Notre Dame Journal of Formal Logic* 7: 103-6. (A report of Asenjo's early work on paraconsistent logic.)
- Batens, D.** (1989) 'Dynamic Dialectical Logics', in G. Priest, R. Routley and J. Norman (eds) *Paraconsistent Logic*, Munich: Philosophia, ch. 6. (The first non-monotonic paraconsistent logic.)
- Brady, R.T.** (1989) 'The Non-Triviality of Dialectical Set Theory', in G. Priest, R. Routley and J. Norman (eds) *Paraconsistent Logic*, Munich: Philosophia, ch. 16. (Includes a proof that naïve set theory with a suitable underlying paraconsistent logic is non-trivial.)
- Costa, N.C.A. da** (1974) 'On the Theory of Inconsistent Formal Systems', *Notre Dame Journal of Formal Logic* 15: 497-510. (A report of da Costa's early work on paraconsistent logic.)
- Costa, N.C.A. da and Marconi, D.** (1989) 'An Overview of Paraconsistent Logic in the 80s', *Journal of Non-Classical Logic* 6: 5-32. (Particularly helpful on work in South America.)
- Jaśkowski, S.** (1969) 'Propositional Calculus for Contradictory Deductive Systems', *Studia Logica* 24: 143-57. (An English translation of the earliest influential paper on paraconsistent logic.)
- Łukasiewicz, J.** (1971) 'The Law of Contradiction in Aristotle', *Review of Metaphysics* 24: 485-509. (An English translation of Łukasiewicz's critique of Aristotle, on the law of non-contradiction.)
- Meyer, R.K. and Mortensen, C.** (1984) 'Inconsistent Models for Relevant Arithmetics', *Journal of Symbolic Logic* 49: 917-29. (On models of inconsistent arithmetics.)
- Priest, G.** (1987) *In Contradiction*, The Hague: Kluwer. (A detailed defence of dialetheism.)
- Priest, G.** (1989) 'Dialectic and Dialetheic', *Science and Society* 53: 388- 415. (A discussion of Hegel, Marx and dialetheism.)
- Priest, G.** (1991a) 'Intensional Paradoxes', *Notre Dame Journal of Formal Logic* 32: 193-211. (The construction of non-trivial theories containing both self-reference and epistemic operators.)
- Priest, G.** (1991b) 'Minimally Inconsistent LP', *Studia Logica* 50: 321-31. (Consistency as a default assumption.)
- Priest, G.** (1984) 'Paraconsistent Logic', in D. Gabbay and F. Guenther (eds) *Handbook of Philosophical Logic*, vol. 2, Dordrecht: Reidel, 2nd edn, forthcoming. (The best reference work on paraconsistent logic currently available. It has a thorough overview of the technical aspects, with some philosophical comment and references to the literature.)
- Priest, G., Routley, R. and Norman, J. (eds)** (1989) *Paraconsistent Logic: Essays on the Inconsistent*, Munich: Philosophia. (This standard reference work includes many interesting papers, a bibliography up to the mid-1980s and editorial essays which go into the topics discussed here at much greater length.)
- Routley, R.** (1977) 'Ultralogic as Universal?', *Relevance Logic Newsletter* 2: 50-90, 138-75; repr. as an appendix in *Exploring Meinong's Jungle, and Beyond*, Canberra: Research School of Social Sciences, Australian National University, 1980. (A statement of Routley's earlier views on paraconsistency.)
- Smiley, T.J.** (1959) 'Entailment and Deducibility', *Proceedings of the Aristotelian Society* 59: 233-54. (The earliest filter paraconsistent logic.)
- Smiley, T.J. and Priest, G.** (1993) 'Can Contradictions be True?', *Proceedings of the Aristotelian Society*, supplementary vol. 67: 17-54. (A debate concerning a number of objections to dialetheism.)

Paradoxes of set and property

Emerging around 1900, the paradoxes of set and property have greatly influenced logic and generated a vast literature. A distinction due to Ramsey in 1926 separates them into two categories: the logical paradoxes and the semantic paradoxes. The logical paradoxes use notions such as set or cardinal number, while the semantic paradoxes employ semantic concepts such as truth or definability. Both often involve self-reference.

The best known logical paradox is Russell's paradox concerning the set S of all sets x such that x is not a member of x . Russell's paradox asks: is S a member of itself? A moment's reflection shows that S is a member of itself if and only if S is not a member of itself - a contradiction.

Russell found this paradox by analysing the paradox of the largest cardinal. The set U of all sets has the largest cardinal number, since every set is a subset of U . But there is a cardinal number greater than that of any given set M , namely the cardinal of the power set, or set of all subsets, of M . Thus the cardinal of the power set of U is greater than that of U , a contradiction. (The paradox of the largest ordinal, discussed below, is similar in structure.)

Among the semantic paradoxes, the best known is the liar paradox, found by the ancient Greeks. A man says that he is lying. Is what he says true or false? Again, either conclusion leads to its opposite. Although this paradox was debated in medieval Europe, its modern interest stems from Russell, who placed it in the context of a whole series of paradoxes, including his own.

1 General comments

Etymologically, a paradox is something 'against' ('para') '[common] opinion' ('dox'). Nowadays it means a claim that seems absurd but has an argument to sustain it. A paradox appears 'paradoxical' when one is uncertain which premise to abandon.

The paradoxes about sets involve the principle of comprehension, which states that, for any concept, there is a class of all those objects for which the concept is true. (We use the terms 'class' and 'set' interchangeably, thus allowing each author discussed below to keep his own terminology.) Historically, the paradoxes about sets are related to the traditional antinomies of the infinite (as expressed, for example, by Galileo), to Zeno's paradoxes and to Kant's antinomies (see §4 below).

2 Burali-Forti: how not to discover a paradox

The paradox of the largest ordinal, often called Burali-Forti's paradox, is as follows. The set of all ordinals is well-ordered and so has an ordinal number, Ω , which is the largest ordinal. But for any ordinal α , there is a larger ordinal $\alpha + 1$. Hence $\Omega + 1$ is larger than Ω , a contradiction.

The early history of this paradox is a comedy of errors. In 1897 Cesare Burali-Forti *almost* discovered it. He failed to do so because he misunderstood Cantor's definition that a linearly ordered set is 'well-ordered' if (a) it contains a first element, (b) every element with a successor has an immediate successor and (c) every set of elements with a successor has an immediate successor. Accidentally omitting (c) from the definition, Burali-Forti introduced a new notion of 'perfectly ordered class' and gave for such classes an argument close to the paradox. But the conclusion he drew was that the order types of such classes are not linearly ordered. He did not think he had discovered a paradox, since nothing that he wrote challenged previous results. Later that year, he realized that he had misconstrued Cantor's definition, but saw no contradiction between his work and Cantor's, since his concerned perfectly ordered classes, not ordinals. For five years, there was no discussion in print about Burali-Forti's article. (This has often been mistakenly denied, as when van Heijenoort stated that Burali-Forti's article 'immediately aroused the interest of the mathematical world' (1967: 104).)

In 1902, after Louis Couturat alerted him to Burali-Forti's work, Russell asserted (wrongly) in print that Burali-Forti claimed that the ordinals are not linearly ordered. Here was the beginning of a genuine paradox, for Russell applied Burali-Forti's argument to the ordinals. Russell's solution was to claim that the class of all ordinals has no ordinal since it is not well-ordered. (Later he rejected this solution.)

In 1903 Russell gave the first clear statement of the paradox of the largest ordinal in his book *The Principles of Mathematics*. Thus Russell was the key figure in its emergence. Alonzo Church erroneously claimed that ‘it was through Burali-Forti’s paper that there first came to general attention the threat to the foundations of mathematics that is constituted by the antinomies’ (1971: 235). It was not Burali-Forti but Russell who first pointed out this threat in print.

3 Cantor: side-stepping the paradoxes

In 1883 Cantor had introduced infinite ordinal numbers (see *Set theory; Cantor’s theorem*). He distinguished sharply between finite ordinals, infinite ordinals and what he called the ‘absolutely infinite’, which included the set of all ordinals. In 1897, when Hilbert asked whether every set can be well-ordered, Cantor answered yes. Cantor’s argument indirectly included the paradox of the largest ordinal and explicitly relied on the absolutely infinite. Yet he did not regard this as paradoxical, since he had previously made the relevant distinction between the infinite and the absolutely infinite. (E.H. Moore independently found this ordinal paradox and wrote about it to Cantor in an unpublished letter of 1898.)

Replying to Hilbert, Cantor made the notion of the absolutely infinite more precise. In print two years earlier, he had defined a set as a ‘gathering together’ of distinct objects. Now he called a set ‘completed’ if it is possible, without contradiction, to think of all its members as gathered together into a whole; any other set, such as that of all ordinals, was ‘absolutely infinite’. In his Paris lecture of 1900, Hilbert affirmed that each of Cantor’s infinite cardinal numbers is consistent, but denied that there exists a set of all cardinals. Nevertheless, Hilbert gave no indication that set theory was threatened.

In 1899 Cantor communicated to Dedekind the proof that every set can be well-ordered (see *Cantor 1991*). Once more, the paradox of the largest ordinal was involved when Cantor showed, as part of the proof, that the collection of all ordinals is absolutely infinite. Now Cantor distinguished between what he called consistent collections and inconsistent collections; only the former were sets. (A ‘consistent’ collection was what he had earlier called ‘completed’.) Writing to Hilbert shortly afterwards, Cantor saw Dedekind’s foundation as threatened, since Dedekind assumed that every well-defined collection was consistent. Cantor had pointed out to Dedekind that the collection of everything thinkable, which Dedekind had used to prove the existence of an infinite set, was inconsistent.

4 Russell: the paradoxes emerge

The paradoxes were not discussed in print until 1903, when they appeared in Russell’s *Principles of Mathematics*. But Russell had found the paradox of the largest cardinal before he learned of Burali-Forti’s article or of Cantor’s unpublished thoughts on the subject. Russell’s own paradox emerged in May 1901 from simplifying his paradox of the largest cardinal.

But the roots of Russell’s paradox go much deeper. In 1896 he had rejected Cantor’s infinite ordinals as ‘impossible and self-contradictory’. During that period Russell proposed various antinomies in the spirit of Kant and Hegel, including the ‘contradiction of relativity’. In 1898 he characterized mathematics as follows:

One pervading contradiction occurs almost... universally. This is the contradiction of a difference between two terms, without a difference in the conceptions applicable to them. I shall call it the contradiction of relativity. This, with addition and the manifold, appear to define the realm of Mathematics.

(1898: 166)

By 1899, Russell was no longer an idealist but a Platonist. Although he ceased propounding antinomies in the style of Kant, he proposed one that originated with Leibniz. ‘Mathematical ideas’, he wrote in his 1899-1900 draft of the *Principles*, ‘are almost all infected with one great contradiction. This is the contradiction of infinity. All antinomies,... so far as they are valid at all, will be found reducible to the antinomy of infinite number’ (1900: 70).

What was this antinomy of infinite number? In 1899 Russell gave it the following form, while distinguishing between

(a) all numbers, (b) the greatest number, (c) the last number. Observe that (b)... means the number applying to the greatest collection. All three are commonly called infinite number, and imply an antinomy, since their being

can be both proved and disproved. (a) the most fundamental: There are many numbers, therefore there is a number of numbers. If this be N , $N + 1$ is also a number, therefore there is no number of numbers.

(1899: 265)

Russell's antinomy of infinite number has the same formal structure as the paradox of the largest cardinal, which he formulated only in late 1900. That paradox can be stated in the following way: the class of all cardinal numbers has a cardinal number; but there is another class which has a larger cardinal number (by Cantor's theorem, the set of subsets of a set with cardinal N has a cardinal larger than N); hence there is no cardinal number of the class of all cardinal numbers. Likewise, the antinomy of infinite number has the same formal structure as his later paradox of the largest ordinal.

In November 1900 Russell redrafted part of the *Principles*. The following passage from that draft manuscript contains the earliest version of his paradox of the largest cardinal, which is closely connected with the antinomy of infinite number.

There is a certain difficulty in regard to the number of numbers, or the number of individuals or of classes. Numbers, individuals, and classes, each form a perfectly definite class, and... every class must have a number. Now the number of individuals must be the absolute maximum of numbers, since every other class is a proper part of this one.... But Cantor has given two proofs... that there is no greatest number.

Soon Russell informed Couturat that Cantor's proofs were erroneous when applied to the class of all classes. Couturat doubted the existence of that class. But Russell vigorously defended it: 'If you grant that there is a contradiction in this concept, then the infinite always remains contradictory' (Moore and Garciadiego 1981: 327).

In May 1901 Russell discovered his paradox by applying Cantor's proof that there is no greatest cardinal to the class of all classes. During that month he wrote the earliest surviving manuscript containing his paradox. It was not expressed in terms of classes but of predicates:

We saw that some predicates can be predicated of themselves. Consider now those... of which this is not the case.... But there is no predicate which attaches to all of them and to no other terms. For this predicate will either be predicable or not predicable of itself.

(1993: 195)

Either case implies the opposite, hence a contradiction. Thus there is no predicate predicable of all and only those predicates not predicable of themselves. Russell concluded that the principle of comprehension was false. A year later, he wrote to Frege about it. While Frege saw that the paradox for predicates did not arise in his system, he was devastated by the class form of Russell's paradox. Frege abandoned the principle of comprehension, which seemed to him the only possible logical foundation for arithmetic.

In 1903 Russell published Russell's paradox, together with the paradoxes of the largest cardinal and the largest ordinal, in the *Principles*. By that date, one or more of these three paradoxes was also known to Dedekind and Hilbert (thanks to Cantor), to Husserl (thanks to Zermelo; see §5 below), to Peano (thanks to Russell) and to E.H. Moore (independently, in a letter to Cantor), but none of them published such a result. Russell's and Frege's reactions to the paradoxes, which they treated as a major challenge, stand in striking contrast to the early reactions of the others mentioned.

Russell's book also generalized his paradox: if R is a relation, suppose there is some w such that, for any x , x has the relation R to w if and only if x does not have the relation R to x ; substituting w for x leads to a contradiction; hence there is no such w (1903: 102). (Russell's paradox results from substituting membership for R .) Later Quine expressed Russell's generalization as follows: nothing can have a relation to all and only those things that do not have it to themselves.

At the end of this book Russell tentatively presented his earliest version of the theory of types as a solution to the paradoxes. Then 'x is a member of x' was meaningless, since if x is a member of u , then x and u had to belong to different types (see §7 below).

5 Hilbert, Bernstein and Grelling: German reaction to the paradoxes

Cantor was not alone in recognizing, before paradoxes were published, that there is no set of all sets. In his 1890 book on the algebra of logic, Ernst Schröder rejected such a universal set, arguing that it leads to a contradiction. Reviewing Schröder's book, Edmund Husserl rejected both Schröder's argument and his conclusion, since both resulted from a failure to distinguish between being a member of a set and being a subset. In 1902 Ernst Zermelo informed Husserl that Schröder was correct in his conclusion, although mistaken in his argument. Zermelo showed that a set which contains each of its subsets as a member is self-contradictory. He did so by stating Russell's paradox, which he had found by 1900, before Russell. Zermelo concluded that there is no set of all sets.

In 1903 Hilbert wrote to Frege that Zermelo had discovered Russell's paradox three years earlier, after Hilbert had informed Zermelo of other paradoxes which Hilbert had found a year before that. (In an 1899 letter to Frege, Hilbert had observed that there is no set of all cardinal numbers, but gave no hint that this was a threatening discovery.) Influenced by Frege rather than Russell, Hilbert first published on the paradoxes in 1905. He rejected the principle of comprehension, regarding it as the source of the paradoxes. These showed that traditional logic did not satisfy the demands of set theory; they would be solved only by an axiomatic development of logic. That year he devoted a long series of lectures to the problem without coming to a solution. Although one often reads that he did not concern himself with logic during the next decade, in fact he repeatedly lectured on the paradoxes. He only reached a solution in 1917 when he adopted Russell's theory of types (see §7 below).

In Germany the first published discussion about the paradox of the largest ordinal occurred in 1905. Felix Bernstein used this paradox while attempting to refute Zermelo's proof (1904) that every set can be well-ordered, arguing that there is indeed a set W of all ordinals (Moore 1982: chap. 2). But in France, Jacques Hadamard rejected Bernstein's argument: 'It is the very existence of the set W that generates a contradiction.... One has the right to form a set only with previously existing objects, and it is easily seen that the definition of W presupposes the opposite' (Hadamard in Moore and Garciadiego 1981: 339).

Several neo-Friesian philosophers commented on the paradoxes, including Alexander Rüstow (1910), in his book on the liar paradox, and Gerhard Hessenberg (1906), who found the notion of set unclear. Hessenberg argued that Russell's paradox was not dangerous for mathematicians, since they did not use a universal set, but that the paradox of the largest ordinal was dangerous.

In 1908 the neo-Friesian philosophers Kurt Grelling and Leonard Nelson discussed those paradoxes, including the predicate form of Russell's paradox. Grelling formulated a related paradox, later called Grelling's paradox: a word is called autological if it applies to itself, and otherwise is called heterological. Thus 'short' is autological and 'long' is heterological. Now 'heterological' is either autological or heterological. But either possibility leads to the opposite.

6 Berry, König and Richard: paradoxes of definability

The first paradoxes of definability emerged during 1904-5 from three distinct sources. Richard's paradox, due to Jules Richard, was the first to be published. Conceived as a response to Hadamard's comments on the paradox of the largest ordinal (see §5 above), but directed at the continuum of real numbers, Richard's paradox was the following: consider in decimal form all the real numbers that can be defined by a finite number of words. These numbers form a denumerable set E , whose members are $a_1, a_2, a_3, \dots$. We define a real number N not in E by letting the n th decimal place of N be one more than the n th place of a_n unless the n th place is 8 or 9 (in which case the n th place of N is 1). Since N differs from the n th place of a_n for each n , then N is not in E . But N is defined by a finite number of words and so is in E , a contradiction.

In 1906 Henri Poincaré stressed the importance of Richard's paradox, and found in it a single explanation for all the paradoxes: objects defined by a vicious circle must be avoided. In Richard's paradox, there is a vicious circle since N is defined by using E but is itself in E . If one restricted E to those numbers defined without reference to E , there would be no contradiction. Yet Poincaré viewed the actual infinite, which he rejected, as being equally the source of the paradoxes.

Although it was the first definability paradox to be published, Richard's was not the first to be discovered. That honour goes to George Berry of Oxford University. In December 1904 Berry sent Russell the following paradox: consider the least ordinal not definable in a finite number of words; this ordinal is itself defined in a finite number of words, a contradiction.

This paradox, which Russell published in 1906 in response to Poincaré's comments on Richard's paradox, is known by the name he gave it: the paradox of the least indefinable ordinal. At that time Russell modified this paradox to give one that did not involve the actual infinite and that he called Berry's paradox, although Berry did not formulate it: the least integer not nameable in fewer than nineteen syllables. But this integer has just been named in eighteen syllables, a contradiction.

A fourth paradox of definability, called the Zermelo-König paradox, was found in 1905 by Julius König, who published it not as a paradox but to refute Zermelo's claim that the continuum of real numbers can be well-ordered. König observed that the set of real numbers which are definable in a finite number of words is denumerable. Since the continuum is not denumerable, there are real numbers x which cannot be defined in a finite number of words. If the continuum can be well-ordered, then there is a least such x , a contradiction. Hence the continuum cannot be well-ordered.

Once Richard's paradox was seen to be similar to König's result, it became clear that they concerned definability and had nothing to do with well-ordering the continuum.

7 Russell: solving the paradoxes by type theory

In 1905 Russell discussed three kinds of solutions to the paradoxes: zigzag theories, theories of limitation of size, and no-classes theories. In a zigzag theory, a property $\phi(x)$ determines the class of all x satisfying $\phi(x)$ when $\phi(x)$ is sufficiently simple; the difficulty with such a theory, Russell noted, was in specifying what made $\phi(x)$ sufficiently simple. A theory of limitation of size, such as that of Cantor or later Zermelo, would state that a property $\phi(x)$ determines the class of all x satisfying $\phi(x)$ if the class is 'limited' in size; the difficulty here was to specify how large an ordinal exists. Finally, a no-classes theory would dispense with assuming the existence of classes and operate directly with properties; although such a theory would preserve classical analysis and geometry, it was unclear how much set theory would survive.

Russell also gave a general form to all paradoxes about classes: suppose that there is a property ϕ and a function f such that if ϕ is satisfied by all members of the class u , then $f(u)$ satisfies ϕ and is not a member of u ; the assumption that there is a class w of all x satisfying ϕ and that $f(w)$ exists leads to the contradiction that $f(w)$ has and does not have the property ϕ . In Russell's paradox, ϕ is ' x is not a member of x ' and $f(u)$ is u ; in the paradox of the largest ordinal, ϕ is ' x is an ordinal' and $f(u)$ is the ordinal number of u .

In the *Principles* (1903) Russell had tentatively presented his earliest version of the theory of types as a solution to the paradoxes. But during 1903 Russell became dissatisfied with his theory of types, and in 1904 adopted a zigzag theory. In 1905 and 1906 he preferred a no-classes theory. But the boundaries between his different theories were fuzzy. In 1906 he stressed that his no-classes theory contained types and was close to his 1902 theory of types.

The emergence of definability paradoxes coincided with a shift in Russell's views. Earlier, he had primarily investigated paradoxes involving classes. In 1906 he was increasingly occupied with the liar paradox. In that context he formulated the vicious circle principle, which any proposed solution to the paradoxes must satisfy: whatever involves a bound variable must not be a possible value of that variable. While accepting Poincaré's analysis that vicious circles caused the paradoxes, Russell vigorously rejected Poincaré's claim that the actual infinite caused them too.

In 1908 Russell published his *ramified* theory of types, which included much of the no-classes theory. Whereas in 1900 Russell had vigorously defended the class of all classes, his theory of types explicitly rejected it (as well as the notions of all propositions, all relations, all definitions and all ordinals). The heart of the theory was the distinction between orders and types, where orders concerned definitions, and the central assumption was his dubious axiom of reducibility, which was used to solve the semantic paradoxes.

Russell's theory was not well received at first, and was rejected by Poincaré in part because of the axiom of reducibility. The *simple* theory of types, which dispensed with this axiom and with definability, was due to Leon Chwistek and Ramsey in the 1920s (see *Theory of types*). In the simple theory, there was no concern with the semantic paradoxes, which could not even be formulated in it.

8 Zermelo: solving the paradoxes by axiomatizing set theory

In 1908 **Zermelo** published his axioms for set theory, without reformulating logic or mentioning the semantic paradoxes. To understand Zermelo's motivation, we must consider two articles he published that year in the same journal, finished within two weeks of each other, and which referred to each other. The second article contained his axiomatization, while the first defended his theorem that every set can be well-ordered. The first also gave a new proof of his theorem, based explicitly on axioms from his second article. The two articles form a single unity, revealing that his axiomatization was motivated as much by a concern to secure his theorem as by the need to solve the paradoxes (see **Moore 1978**).

The key to Zermelo's axiomatization was his axiom of separation, which restricted the principle of comprehension to subsets of any previously existing set. To preserve set theory, he included axioms which generated sets from previously existing sets, especially the power set and union axioms (see **Set theory, different systems of**).

But Zermelo's approach seemed arbitrary until he introduced the cumulative type hierarchy of sets in 1930. This hierarchy began with the empty set as the lowest level, and any level S was followed by a next level, the power set of S . The levels were indexed by ordinals, and when the index was a limit ordinal, the level was the union of all previous levels. The hierarchy of levels was cumulative since a member of any level belonged to all following levels, by contrast with Russell's types, which were disjoint.

Today Zermelo's hierarchy, as embodied in the Zermelo-Fraenkel axioms formulated in first-order logic, is generally accepted by mathematicians as the proper solution to the logical paradoxes. Among philosophers, the semantic paradoxes still lack such a generally accepted solution.

9 Brouwer, Weyl and Skolem: attacking set theory with paradoxes

Opponents of set theory repeatedly used the paradoxes as a weapon against it. This began with Poincaré in 1906 and continued with Brouwer in 1907 (see **Intuitionism**). Both saw intuition as central, the axiomatization of logic as suspect, and the natural numbers as fundamental. Both engaged in later attacks on axiomatization and set theory. A decade later Hermann **Weyl** continued the attack on analysis and set theory, using Grelling's and Richard's paradoxes.

In 1923 Thoralf Skolem used a different paradox to argue that set-theoretic concepts are relative, not absolute. Skolem's paradox states that there is a countable model for axiomatic set theory, although there exist uncountable sets. The solution distinguishes between levels. There exists some set M which is uncountable within the model, where there is no function mapping M onto the natural numbers; but outside the model there is such a function, and so, from outside, all the sets in the model are seen as countable.

10 Zermelo, von Neumann and Gödel: turning paradoxes into theorems

From a mathematical perspective, what is most striking is how paradoxes have repeatedly been turned into theorems. We consider four examples.

For Zermelo, Russell's paradox became the theorem of his axiomatic set theory that there is no set of all sets (1908).

For von **Neumann**, the paradox of the largest ordinal became the theorem of his axiomatic set theory (1925) that every set can be well-ordered. The proof was based on the axiom that every proper class (and no set) can be mapped one-to-one onto the class of all sets. Thus proper classes, which were 'too big' to be sets, were rehabilitated by no longer being considered to be contradictory, as they had been for Cantor and Zermelo. In effect, von Neumann made precise Cantor's distinction between consistent and inconsistent collections.

Gödel's incompleteness theorem, which states that any sufficiently rich formal system contains propositions that cannot be proved or refuted in the system (see **Gödel's theorems**), was obtained, in part, by applying Richard's paradox to the notion of provability (1931).

Finally, Tarski's theorem on the indefinability of truth (1933) had a close connection with the liar paradox. Tarski saw the universality of natural languages as the source of the semantic paradoxes, since this universality allows truth to be defined in such languages (see **Tarski's definition of truth**).

See also: **Logical and mathematical terms, glossary of**; **Semantic paradoxes and theories of truth**

References and further reading

- Bentham, J.F. van** (1978) 'Four Paradoxes', *Journal of Philosophical Logic* 7: 49-72. (Analyses the interrelations between the paradox of the largest cardinal, Russell's paradox, Curry's paradox about self-reference without negation, and Löb's version of the liar paradox.)
- Beth, E.W.** (1964) *The Foundations of Mathematics*, Amsterdam: North Holland. (Part 6 has an older, but still useful, summary of the paradoxes.)
- Cantor, G.** (1991) *Briefe (Letters)*, ed. H. Meschkowski and W. Nilson, Berlin: Springer. (Includes Cantor's 1899 letters to Dedekind and 1897-1900 letters to Hilbert relating to the paradoxes.)
- Church, A.** (1971) 'Logic, History of, IV: Modern Logic', in *Encyclopaedia Britannica* vol. 14: 231-7. (A general discussion, including the paradoxes.)
- Fraenkel, A.A., Bar-Hillel, Y. and Levy, A.** (1973) *Foundations of Set Theory*, Amsterdam: North Holland. (A careful analysis of the various solutions of the logical paradoxes, including axiomatic set theory, the theory of types, and metamathematics.)
- Garciadiego, A.** (1985) 'The Emergence of the Non-Logical Paradoxes of the Theory of Sets, 1903- 1908', *Historia Mathematica* 12: 337-51. (The early history of the semantic paradoxes.)
- Greiling, K. and Nelson, L.** (1908) 'Bemerkungen zu den Paradoxieen von Russell und Burali-Forti' (Remarks on the Paradoxes of Russell and Burali-Forti), *Abhandlungen der Fries'schen Schule*, new series, 2 (3): 301-34. (Greiling's paradox.)
- Heijenoort, J. van** (ed.) (1967) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press. (Translates relevant articles by Burali-Forti, Gödel, König, Richard, von Neumann and Zermelo. Includes Russell's 1902 letter announcing his paradox to Frege and mentioning that he had previously sent the paradox to Peano but received no reply.)
- Hessenberg, G.** (1906) *Grundbegriffe der Mengenlehre (Fundamental Concepts of Set Theory)*, Göttingen: Vandenhoeck & Ruprecht. (The earliest textbook of set theory.)
- Hughes, P. and Brecht, G.** (1975) *Vicious Circles and Infinity: A Panoply of Paradoxes*, Garden City, NY: Doubleday. (A popular anthology of many paradoxes.)
- Moore, G.H.** (1978) 'The Origins of Zermelo's Axiomatization of Set Theory', *Journal of Philosophical Logic* 7: 307-29. (The role of the axiom of choice and the paradoxes in axiomatization.)
- Moore, G.H.** (1982) *Zermelo's Axiom of Choice: Its Origins, Development, and Influence*, New York: Springer. (The second chapter indicates how the paradoxes were intertwined with Zermelo's proof that every set can be well-ordered.)
- Moore, G.H.** (1988) 'The Roots of Russell's Paradox', *Russell: The Journal of the Bertrand Russell Archives*, new series, 8: 146-56. (The connection between Russell's paradox and his antinomy of infinite number.)
- Moore, G.H. and Garciadiego, A.** (1981) 'Burali-Forti's Paradox: A Reappraisal of its Origins', *Historia Mathematica* 8: 319-50. (Detailed historical treatment of this paradox.)
- Peckhaus, V.** (1990) *Hilbertprogramm und kritische Philosophie: das Göttinger Modell interdisziplinärer Zusammenhänge zwischen Mathematik und Philosophie (Hilbert's Programme and Critical Philosophy)*, Göttingen: Vandenhoeck & Ruprecht. (Includes reaction to the paradoxes by the neo-Friesian school.)
- Rang, B. and Thomas, W.** (1981) 'Zermelo's Discovery of the "Russell Paradox"', *Historia Mathematica* 8: 15-22. (How Zermelo found Russell's paradox, prior to Russell, and discussed it with Husserl.)
- Russell, B.A.W.** (1898) 'An Analysis of Mathematical Reasoning, Being an Inquiry into the Subject Matter, the Fundamental Conceptions, and the Necessary Postulates of Mathematics', in *The Collected Papers of Bertrand Russell*, vol. 2, *Philosophical Papers 1896-99*, ed. N. Griffin and A.C. Lewis, London: Routledge, 1990, 155-242. (Includes the contradiction of relativity.)
- Russell, B.A.W.** (1899) 'The Fundamental Ideas and Axioms of Mathematics', in *The Collected Papers of Bertrand Russell*, vol. 2, *Philosophical Papers 1896-99*, ed. N. Griffin and A.C. Lewis, London: Routledge, 1990, 261-305. (Includes the antinomy of infinite number.)
- Russell, B.A.W.** (1903) *The Principles of Mathematics*, Cambridge: Cambridge University Press; 2nd edn, London: Allen & Unwin, 1937; repr. London: Routledge, 1992. (First publication of Russell's paradox, the paradox of the largest cardinal and the paradox of the largest ordinal.)
- Russell, B.A.W.** (1906) 'Les paradoxes de la logique', *Revue de métaphysique et de morale* 14: 627-50; trans. 'On "Insolubilia" and their Solution by Symbolic Logic', in *Essays in Analysis*, ed. D. Lackey, London: Allen & Unwin, 1973. (Analysis of the paradoxes, including the semantic ones.)
- Russell, B.A.W.** (1993) *The Collected Papers of Bertrand Russell*, vol. 3, *Toward the 'Principles of*

Mathematics '1900-02, ed. G.H. Moore, London and New York: Routledge. (See 'The Principles of Mathematics (Draft of 1899-1900)' (3-180) on the antinomy of infinite number; and 'Part I of the *Principles*, Draft of 1901' (181-208) for the earliest extant version of Russell's paradox.)

Rüstow, A. (1910) *Der Lügner (The Liar)*, Leipzig: Teubner. (The liar paradox.)

Zermelo, E. (1904) 'Beweis, daß jede Menge wohlgeordnet werden kann', *Mathematische Annalen* 59: 514-16; trans. S. Bauer-Mengelberg, 'Proof that Every Set can be Well-Ordered', in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967, 139-41. (Zermelo's proof of the well-ordering theorem.)

Polish logic

The term 'Polish logic' was coined by McCall to signal the important contributions to modern logic by logicians from Poland between the wars. There were several centres of research, of which the Warsaw school, which grew out of the earlier Lwów-Warsaw philosophical movement, was the most significant. Its development was closely connected with the Warsaw school of mathematics, which gave it its characteristic mathematical bent.

Polish logic took as its point of departure the main trends in logical research of the time and it has influenced both subsequent logical research and subsequent work in the Western analytic tradition of philosophy. Its chief contributions were: (1) an enrichment of existing logical theory (including work on Boolean algebras, the sentential calculus, set theory, the theory of types); (2) new logical theories (for example, Leśniewski's systems, Łukasiewicz's many-valued logics, Tarski's theory of truth, theory of the consequence operation and the calculus of systems); (3) new methods and tools as well as improvements of existing methods (for example, the matrix method of constructing sentential calculi, axiomatizability of logical matrices, algebraic and topological interpretations of deductive systems, permutation models for set theory, the application of quantifier elimination to decidability and definability problems); and (4) the application of formal methods to the study of the history of logic, resulting in a new understanding of the logics of Aristotle, the Stoics and the medievals.

1 Historical introduction

In 1918, after 120 years of subjugation by foreign powers, Poland reappeared on the map of Europe as a free and independent nation-state. In 1939 the outbreak of the Second World War reversed these brief gains, bringing down the curtain on the Second Republic and inflicting massive damage on its economy, culture and learning. The main centres of logic research between these dates were Cracow, Lwów, Poznan, Wilno and, most importantly, Warsaw. The Warsaw school of logic, founded by Łukasiewicz and Leśniewski, began as an intellectual offshoot of the Lwów-Warsaw philosophical movement, but quickly eclipsed it in both quantity and quality of research.

The development of the Warsaw school of logic was closely connected with that of the Warsaw school of mathematics. In 1918 one could already speak of Warsaw as a fairly strong centre of research in set theory and topology under the direction of Janiszewski, Mazurkiewicz and Sierpiński. The two schools shared organizational structures, swapped directors and collaborated on many academic initiatives. This accelerated the development of both schools and deepened their research in key areas. In later years Kuratowski, Lindenbaum, Tarski and finally Mostowski made significant contributions to both schools.

2 Łukasiewicz: sentential calculi and history of logic

Łukasiewicz's more philosophical research culminated in his study of the principle of contradiction in Aristotle (1910), which was an important influence on the development of logic in Poland. After his discovery of a 'non-Aristotelian' many-valued logic (1918-20; see Łukasiewicz, J. §3), he turned to the study of the logic of propositions - in particular, to the sentential calculus, which he believed to be the simplest deductive discipline central to all scientific reasoning, and around which he built a research programme which served as a focus for research work on propositional logics throughout Poland (see Łukasiewicz 1941).

Łukasiewicz's programme can be seen as divided into two periods. The first, up to 1930, is characterized by conceptual and foundational achievements (see Łukasiewicz and Tarski 1930; Łukasiewicz 1929).

Notable among the conceptual achievements are a definition of a recursive grammar for propositional languages (including propositional quantifiers); the now famous bracket-free notation known as Polish notation; a precise definition of the syntactic operations of substitution and replacement; a new axiom system for classical propositional logic together with a new completeness proof for this system; formulations of several systems of many-valued logic (see *Many-valued logics, philosophical issues in*); a rigorous distinction between considerations of syntax and proof theory, on the one hand, and metalogical and semantic considerations on the other; and the invention of the matrix method of constructing sentential calculi.

Among the foundational achievements are various technical results in the search for single-sentence axiomatizations (notably by Tarski), Wajsberg's discovery of a criterion for finite axiomatizability, Lindenbaum's

theorem on the existence of adequate matrices for arbitrary systems of propositional logic closed under substitution, and pioneering studies of partial calculi (for example, the pure implicational and the pure equivalential fragments of classical sentential logic) and of sentential calculi extended by propositional quantifiers and variable functors of higher order.

The second period, from 1931 on, is characterized by the expansion of the programme to cover a greater number of propositional calculi, and its enrichment with new methods drawn from algebra and topology.

Noteworthy results of this period include a complete algebraic semantics for Lewis' S5 (see Wajsberg 1933), Jaśkowski's invention - independently of Gentzen - of the idea of a system of natural deduction (see Jaśkowski 1934; *Natural deduction, tableau and sequent systems* §1), a complete algebraic semantics for intuitionistic logic (Jaśkowski 1936), a topological interpretation of certain sentential calculi, notably the intuitionist variant (Tarski 1956), and a proof of the separability of the intuitionist connectives (Wajsberg 1938).

Still, a great deal of research in this period continued to be devoted to finite logical matrices. Wajsberg (1935), Sobociński (1936) and Słupecki (1939a) worked on solutions to the axiomatizability problem for many-valued logics. Słupecki (1939b) found a criterion for functional completeness of finite algebras.

Łukasiewicz also founded a programme of historical research whose goals were both a thorough exegesis and an analytical reconstruction of ancient and medieval logic. The source of this programme was his own book on Aristotle (1910), which distinguished ontological, logical and psychological formulations of the principle of contradiction. Łukasiewicz himself traced the historical roots of propositional logic to Stoic dialectic and countered prevailing views of the Middle Ages by shedding new light on the importance of the achievements of Petrus Hispanus and Duns Scotus (1934); he provided an axiomatic reformulation of Aristotle's syllogistic (1929, 1939, 1951; *Logic, ancient* §7). In the course of these investigations, Łukasiewicz introduced a new formal idea: the concept of a 'rejection system' - a Hilbert-type system for disproving rather than proving formulas - which he intended as a syntactic counterpart to the semantic device of refutation. This idea was subsequently elaborated by Słupecki (1949).

Under Łukasiewicz's influence a number of scholars outside Warsaw undertook related research. In Wilno, Czeżowski (1936) studied the treatment of modal sentences in Aristotle and Korcik (1937) provided an interpretation of some of Aristotle's writings in terms of a theory of the conversion of assertoric sentences. Bocheński (1938) wrote a history of the treatment of modal sentences in ancient and medieval times. In Poznań, Jordan (1937) studied the use of the axiomatic method from Plato and Eudoxus to Euclid's *Elements*. In Cracow, Salamucha analysed Saint Thomas' *ex motu* proof of the existence of God (1934), showed that William of Ockham had anticipated large parts of classical propositional calculus (1935) and wrote a scholarly treatise on medieval antinomies (1937).

Łukasiewicz's programme revealed that ancient and medieval logicians had considered 'formal' problems similar in character to some of the problems addressed by contemporary formal logic. This prompted a new interest in the subject and eventually led to a critical re-evaluation of large parts of the history of logic (see Scholz 1931; Łukasiewicz 1951; Mates 1953; Bocheński 1951, 1956; Kotarbiński 1957).

3 Leśniewski: foundations of mathematics

The main goal of Leśniewski was to provide a foundation for mathematics that would yield a real understanding of the antinomies and not simply avoid them. He did not accept the principles and assumptions of Cantorian set theory, regarding them as artificial and counterintuitive. He found the notion of an empty set an affront to reason. He was suspicious of *Principia Mathematica* on the grounds that its formalism was careless, ambiguous and did not pay sufficient attention to the use/mention distinction. He was an uncompromising critic of formalism for formalism's sake, and sought to ground mathematics in extensionalism, constructive nominalism and two-valued logic, insisting that only these principles would assure that it expressed meaningful, true sentences about the real world. He was, in the words of Tarski, the first person to become fully aware that a language which contains its own semantics and within which the usual logical laws hold must inevitably fall prey to liar-type paradoxes and hence be inconsistent.

The cornerstones of Leśniewski's work are *Mereology* (1927-31, outlined in 1916), *Ontology* (1920), his theory of

semantic categories (1929) and *Protothetic* (1929) (see [Leśniewski 1929](#)). The notion of a semantic category, due originally to Husserl, plays a role in Leśniewski's construction of formal theories similar to that played by the notion of type in *Principia Mathematica*. Ajdukiewicz later reformulated the theory of semantic categories in the form of a calculus (see [Ajdukiewicz, K. §5](#); [Ajdukiewicz 1978, article 7](#)).

Constructive nominalism had important metalogical consequences for Leśniewski, confronting him with problems of how to describe a system, its language, axioms and rules (especially rules of definition). Notoriously, he clung to the belief that at any particular moment a logical system contains only as many theorems as have been proved up to that point in time. This led him to an extreme constructivism in the theory of definitions. Definitions in his system were 'creative', that is, after introducing definitions of new terms, the system was, in his sense, stronger than before. His definitions can thus not be regarded as mere abbreviational conveniences.

Tarski, in his 1924 doctoral dissertation under Leśniewski, investigated a fragment of *Protothetic* which might be called the second-order propositional calculus. He showed that in this system all sentential functors can be defined in terms of the equivalence connective and universal quantifier; and that one can prove counterparts to all the basic metalogical laws for classical propositional calculus. These results supported Leśniewski's conviction that metatheoretical considerations outside the system were dispensable.

Mereology influenced Tarski in his papers on solid geometry and Boolean algebras, and also found application in Woodger and Tarski's work on the axiomatic foundations of biology (see [Woodger 1937](#)). Sobociński (1934) published successive simplifications of the axiom system for *Ontology*. Kotarbiński's philosophy of reism, and his classic *Elements* (1929), owe much to the ideas of Leśniewski.

Leśniewski's influence spread chiefly through lectures and seminars. He was unquestionably one of the founding fathers of logic in Poland between the wars. Only after the Second World War, through some of his students' publications and a monograph by Luschei (1962), did the wider community take an interest in his achievements (see [Srzednicki, Rickey and Czelakowski 1984](#)). Some notes from his Warsaw lectures were collected and published in 1988.

4 Tarski: semantics and metamathematics

The role of [Tarski](#) in the development of Polish logic was in every way exceptional. Up to 1939 he published about sixty research papers covering various aspects of logic, set theory, measure theory (including the famous Banach-Tarski theorem on the 'paradoxical' decomposition of the sphere), axiomatic and structural aspects of Boolean algebras, truth theory and other areas of semantics (see [Tarski 1956](#)).

Tarski's contributions in pure logic were outstanding, and his work in formal or 'scientific' semantics and metamathematics helped shape Polish logic. His signal achievement was to establish that, for languages with specifiable structure, semantic notions such as satisfiability, truth, consequence and definability could be treated mathematically, and that this was fruitful (see [Tarski's definition of truth](#); [Consequence, conceptions of](#); [Model theory](#)).

A metamathematical theme was present in Polish logic from the early 1920s, beginning with Ajdukiewicz (1921). Tarski's efforts allowed investigations of the sentential calculi to be carried out within an explicitly metamathematical framework. Tarski generalized this framework to a mathematical theory of two primitive concepts (sentence and consequence) which he called 'the methodology of the deductive sciences'. Within this theory he was able to provide a conceptual apparatus for investigating deductive systems. (Lindenbaum contributed much to this work, including the widely known Lindenbaum maximality lemma.) Tarski later enlarged his metamathematics to encompass the calculus of systems and the concepts of ω -consistency and ω -completeness. These contributions were of a foundational nature, explicitly concerned with metatheory *per se*, and consciously framed in a meta-metalanguage. (See [Tarski 1956](#).)

One of Tarski's best-known discoveries is the 1930 theorem on the eliminability of quantifiers for the theory of real closed fields, together with its consequences for the decidability of the arithmetic of real numbers and elementary geometry, and the definability of sets of reals. To this category also belongs Tarski and Mostowski's work of the late 1930s on the decidability of the theory of well-orderings.

A recurring motif in Tarski's work was the idea of algebraization. He believed that by abstracting an algebraic

calculus from a theory, the tools developed for the theory could be extended to other disciplines. This idea exerted a far-reaching influence on the subsequent course of logical investigations in Poland and elsewhere.

Largely through Tarski's work on truth, Polish logic also influenced philosophy. He spoke out against allowing one's repertoire of methods to be constrained by philosophical 'isms' such as formalism, logicism, constructivism or intuitionism (though his own scientism was exempt), and urged free use of all the resources of set theory, including transfinite techniques. He never accepted Wittgenstein's thesis that there is only one language of philosophical and scientific discourse, and persuaded some members of the Vienna Circle (most notably Carnap) to set it aside as well (see [Tarski 1992](#)).

5 Other work

Four recurring themes characterized the work of the Warsaw set theorists: (1) the axiom of choice, its equivalents and various weakenings, and its role within and beyond set theory; (2) the continuum hypothesis, its role in mathematics and its relation to the axiom of choice; (3) the problem of the independence of the axiom of choice from other principles of set theory; and (4) the arithmetic of cardinal and ordinal numbers (see [Axiom of choice](#); [Set theory](#)).

Sierpiński's paper ([1918](#)) was the first major contribution to (1). He argued that before an informed decision could be made concerning whether to adopt the axiom of choice, one should first investigate its deductive strength. In 1924, Tarski discovered that several statements in the arithmetic of infinite cardinals were equivalent to the axiom of choice and also systematized the theory of finite sets and showed that it was necessary to assume the axiom of choice to prove the equivalence of various definitions of finiteness (see [Tarski 1986: vol. 1, 39-48, 65-117](#)). Lindenbaum was the first to pose the question 'What is the logical relation between the continuum hypothesis and the axiom of choice?'. These and related results were summarized in a 1926 paper by Lindenbaum and Tarski (see [Tarski 1986: vol. 1, 171-204](#)), where the two authors famously stated, without proof, that the generalized continuum hypothesis implies the axiom of choice ([188, theorem 94](#)). In 1934, Sierpiński published a compendium of principles known to be consequences of the generalized continuum hypothesis, but it was not until 1947 that he was able to supply the missing proof ([1976: 485-8](#)).

In 1938 Mostowski investigated the independence of definitions of finiteness in a system of type theory by means of the method of relativization of quantifiers (see [Mostowski 1979: vol. 2, 18-67](#)). During the years 1935-8 he and Lindenbaum also refined and extended Fraenkel's permutation models to investigate the independence of the axiom of choice from the ordering principle and various weaker forms of the axiom ([vol. 1, 290-338; vol. 2, 70-4](#)).

In Cracow, Sleszyński published a survey of the development of formal logic from Aristotle to Russell and Whitehead. Especially valuable is his treatment of nineteenth-century mathematical logic ([1925-9](#)). W. Wilkosz wrote textbooks on modern set theory and axiomatic foundations of number theories.

Chwistek developed a system for the foundation of logic and mathematics. Starting from several improvements of the ramified theory of types (see [Theory of types](#)), he arrived at a purely constructivist theory without the axioms of reducibility and extensionality and without impredicative definitions ([1924-5](#)). Later, with his pupils J. Herzberg and W. Hetper, he outlined a formal theory of expressions called 'rational metamathematics' ([1935](#)). Chwistek's work in logic was strongly influenced by his philosophical beliefs, which set him apart from the mainstream of Polish logic ([1961-3](#)).

Pepis ([1937, 1938](#)) investigated the decision problem for the predicate calculus; more specifically, he dealt with the problem of reducing the question of the satisfiability of a logical formula from the general case to that of a formula having a special normal form.

6 Post-war developments

The Second World War caused the deaths of many logicians, including Lindenbaum and his wife J. Hosiasson-Lindenbaum (who worked on inductive logic), Herzberg, Hetper, Pepis, Salamucha, Z. Schmierer and Wajsberg. Others were forced to emigrate: Bocheński, H. Hiż, Jordan, Łukasiewicz, H. Mehlberg, Sobociński and Tarski. Manuscripts were lost or destroyed, among them Łukasiewicz's manuscript for a book on syllogistic, Mostowski's notes on the proof of quantifier elimination for well-orderings, and Leśniewski's *Nachlaß*. Galley

proofs of a new journal, *Collectanea Logica*, were destroyed in 1939 and the journal never appeared.

Post-war Polish logic picked up where pre-war work had left off. In mathematical logic and the foundations of mathematics, the main figure was Mostowski (1979), whose contributions were wide-ranging, along with A. Grzegorzczuk, A. Erenfeucht, J. Łoś, C. Ryll-Nardzewski and R. Suszko. Łoś is best known for his 1954 notion of categoricity-in-power, his 1955 lemma on ultraproducts and generally for his work on model theory (parts of which were carried out in collaboration with Suszko). Ryll-Nardzewski contributed to model theory (omitting types and ω -categoricity) and arithmetic (showing first-order Peano arithmetic is not finitely axiomatizable). Together, Łoś and Ryll-Nardzewski shed light on the logical importance of Tychonoff's compactness theorem and the prime ideal theorem. Joint papers by Grzegorzczuk, Mostowski and Ryll-Nardzewski (see Mostowski 1979: vol. 1, 371-89; vol. 2, 427-31) are important for our understanding of natural numbers and ω -logic. Mostowski's expository essays and surveys have influenced the wider philosophical community (see Mostowski 1952, 1955, 1965).

Tarski's ideas of the algebraization of logic and of metamathematics were elaborated in numerous papers by Rasiowa and Sikorski and culminated in a classic work, *The Mathematics of Metamathematics* (1963).

In 1958 Łoś and Suszko proposed applying Tarski's theory of the consequence operation to investigations of abstract sentential calculi. This became a research paradigm of post-war Polish logic, drawing on concepts and tools from contemporary model theory and universal algebra. Its principal contributors were J. Czelakowski, W. Dziobak, W.A. Pogorzelski, T. Prucnal, Suszko, P. Wojtylak, R. Wójcicki and A. Wroński. Wójcicki (1987) summarizes the most important results of this programme.

The following logical journals were established in Poland after the Second World War: *Studia Logica* (in 1953), *Bulletin of the Section of Logic* (in 1972), *Reports on Mathematical Logic* (in 1973) and *Logic and Logical Philosophy* (in 1993). An account of post-war Polish logic can be found in Wójcicki (1997) and that of the foundations of mathematics in Marek (1977).

7 Polish notation

In Polish logical notation all operators are placed before the variables over which they have scope.

Np	Negation ($\neg$)
Kpq	Conjunction ($\wedge$)
Apq	Weak or inclusive disjunction ($\vee$)
Jpq	Strong or exclusive disjunction, or non-equivalence ($\neq$)
Cpq	Conditional ($\rightarrow$)
Epq	Biconditional ($\leftrightarrow$)
Dpq	Sheffer stroke or non-conjunction ($ $)
Πpq	Universal quantification ($\forall$) (of q with respect to the sentential variable p)
Σpq	Existential quantification ($\exists$) (of q with respect to the sentential variable p)
ΠxFx	Universal quantification (of Fx with respect to the individual variable x)
ΣxFx	Existential quantification (of Fx with respect to the individual variable x)

Because we know for each operator whether it is one-place or two-place, we know how many of the following letters fall within its scope. So parentheses are not required to avoid ambiguity or to show association. For example, $p \rightarrow (q \rightarrow r)$ is written $CpCqr$ and $(p \rightarrow q) \rightarrow r$ is $CCpqr$; and

$$((p \wedge q) \rightarrow \neg r) \rightarrow (r \rightarrow (\neg p \vee \neg q))$$

is *CCKpqNrCrANpNq*.

The parenthesis-free notation, later called Polish notation, was invented by Łukasiewicz in 1924 and published (1929) in application to the sentential calculus, the sentential calculus with quantifiers and Aristotle's syllogistic. (See also Łukasiewicz and Tarski 1930: 39, note 2; 54-.) The notation is very convenient in theoretical considerations; sometimes it is also used in textbooks (see, for example, Prior 1955).

See also: [Logical and mathematical terms, glossary of](#)

JAN ZYGMUNT

References and further reading

- Ajdukiewicz, K.** (1921) *Z metodologii nauk dedukcyjnych*, Lwów: Nakładem Polskiego Towarzystwa Filozoficznego; trans. J. Giedymin, 'From the Methodology of Deductive Sciences', *Studia Logica* 19: 9-46, 1966. (Referred to in §4.)
- Ajdukiewicz, K.** (1978) *The Scientific-World Perspective and other Essays 1931-1963*, ed. J. Giedymin, Dordrecht: Reidel. (Twenty-two papers in English translation with an introduction by the editor comparing Ajdukiewicz's radical conventionalism to the conventionalism of Poincaré and LeRoy; referred to in §3.)
- Bocheński, I.M.** (1938) *Z historii zdańmodalnych (From the History of Modal Sentences)*, Lwów: Wydawnictwo OO. Dominikanów. (Referred to in §2.)
- Bocheński, I.M.** (1951) *Ancient Formal Logic*, Amsterdam: North Holland. (Referred to in §2.)
- Bocheński, I.M.** (1956) *Formale Logik*, Freiburg and Munich: Alber; trans. I. Thomas, *A History of Formal Logic*, Notre Dame, IN: University of Notre Dame Press, 1961; repr. New York: Chelsea, 1970. (Referred to in §2.)
- Chwistek, L.** (1924-5) 'The Theory of Constructive Types (Principles of Logic and Mathematics)', *Annales de la Société Polonaise de Mathématique* 2: 9-48, 3: 91-141. (Referred to in §5.)
- Chwistek, L.** (1935) *Granice nauki: zarys logiki i metodologii nauk ścisłych*, Warsaw and Lwów: Książnica Atlas; revised edn trans. H.C. Brodie and A.P. Coleman, *The Limits of Science: Outline of Logic and of the Methodology of the Exact Sciences*, London: Routledge & Kegan Paul, 1948. (Referred to in §5.)
- Chwistek, L.** (1961-3) *Pisma filozoficzne i logiczne (Philosophical and Logical Writings)*, ed. K. Pasenkiewicz, Warsaw: PWN, 2 vols. (Almost all of Chwistek's works written in Polish: in volume 1, his philosophical, methodological and aesthetic writings; in volume 2, his logical and mathematical ones (together with *Granice nauki*). Referred to in §5.)
- Czeżowski, T.** (1936) 'Arystotelesa teoria zdańmodalnych' (Aristotle's Theory of Modal Propositions), *Przegląd Filozoficzny* 39 (3): 232-41. (Referred to in §2.)
- Jaśkowski, S.** (1934) *On the Rules of Supposition in Formal Logic*, Warsaw: Nakładem Seminarium Filozoficznego Wydziału Matematyczno-Przyrodniczego Uniwersytetu Warszawskiego; repr. with notational changes in S. McCall (ed.) *Polish Logic 1920-1939*, Oxford: Clarendon Press, 1967. (Referred to in §2.)
- Jaśkowski, S.** (1936) 'Recherches sur le système de la logique intuitioniste', in *Actes du Congrès International de Philosophie Scientifique, Paris 1935*, vol. 6, *Actualités Scientifiques et Industrielles*, Paris: Hermann, 58-61; trans. in S. McCall (ed.) *Polish Logic 1920-1939*, Oxford: Clarendon Press, 1967. (Referred to in §2. For a detailed elaboration of Jaśkowski's semantics for intuitionist logic see [Surma \(1971\)](#).)
- Jordan, Z.** (1937) *O matematycznych podstawach systemu Platona: z historii racjonalizmu (Mathematical Foundations of Plato's System: On the History of Rationalism)*, with a French summary, Poznań: Nakładem Poznańskiego Towarzystwa Przyjaciół Nauk. (Referred to in §2.)
- Jordan, Z.** (1945) *The Development of Mathematical Logic and of Logical Positivism in Poland Between the Two Wars*, Oxford: Oxford University Press; partly repr. in S. McCall (ed.) *Polish Logic 1920-1939*, Oxford: Clarendon Press, 1967, 346-406. (Succinct and readable, but without technical details; lacks historical perspective.)
- Korciak, A.** (1937) *Teoria konwersji zdań asertorycznych u Arystotelesa w świetle teorii dedukcji: studium historyczno-krytyczne (The Theory of Conversion of Assertoric Propositions in Aristotle in the Light of the Theory of Deduction: Critical and Historical Study)*, Wilno: Wydawnictwo Studiów Teologicznych. (Referred to in §2.)
- Kotarbiński, T.** (1929) *Elementy teorii poznania, logiki formalnej i metodologii nauk (Elements of the Theory of Knowledge, Formal Logic and the Methodology of Sciences)*, Lwów: Ossolineum; 2nd revised edn, Warsaw: PWN, 1961; 2nd edn trans. O. Wojtasiewicz, *Gnosiology: The Scientific Approach to the Theory of Knowledge*, Oxford: Pergamon Press, and Wrocław: Ossolineum, 1966. (Referred to in §3.)

- Kotarbiński, T.** (1957) *Wykłady z dziejów logiki (Lectures on the History of Logic)*, Wrocław and Łódź: Ossolineum; 2nd edn, Warsaw: PWN, 1985.(Referred to in §2.)
- Leśniewski, S.** (1988) *S. Leśniewski's Lecture Notes in Logic*, ed. J.T.J. Srzednicki and Z. Stachniak, Dordrecht: Kluwer.(Notes from lectures and seminars given by Leśniewski in the period 1919-39. They cover topics ranging from his own systems to Peano arithmetic and Whitehead's theory of events. Referred to in §3.)
- Leśniewski, S.** (1992) *Collected Works*, ed. S.J. Surma, J.T. Srzednicki and D.I. Barnett, with an annotated bibliography by V.F. Rickey, Warsaw: PWN, and Dordrecht: Kluwer, 2 vols.(English translation of all Leśniewski's published works. Flawed in some respects; see a review in *History and Philosophy of Logic* 15: 227-35, 1994.)
- Łukasiewicz, J.** (1910) *O zasadzie sprzeczności u Arystotelesa: studium krytyczne (On the Principle of Contradiction in Aristotle: A Critical Study)*, Cracow: Polska Akademia Umiejętności; 2nd edn with an introduction and notes by J. Woleński, Warsaw: PWN, 1987.(Referred to in §2.)
- Łukasiewicz, J.** (1929) *Elementy logiki matematycznej*, ed. M. Presburger, Warsaw: PWN; 2nd revised and abridged edn, ed. J. Śłupecki, Warsaw: PWN; trans. *Elements of Mathematical Logic*, Oxford: Pergamon Press, and New York: Macmillan, 1963.(Referred to in §2.)
- Łukasiewicz, J.** (1934) 'Z historii logiki zdań', *Przegląd Filozoficzny* 37: 417-37; German trans. by the author, 'Zur Geschichte der Aussagenlogik', *Erkenntnis* 5: 111-31, 1935; trans. On the History of the Logic of Propositions, in *Selected Works*, ed. L. Borkowski, Amsterdam: North Holland, 1970; and in S. McCall (ed.) *Polish Logic 1920-1939*, Oxford: Clarendon Press, 1967.(According to H. Scholz, the most interesting thirty pages ever written on the history of logic. Referred to in §2.)
- Łukasiewicz, J.** (1939) 'O sylogistyce Arystotelesa' (On Aristotle's Syllogistic), *Sprawozdania Polskiej Akademii Umiejętności* 44: 220-7.(Referred to in §2.)
- Łukasiewicz, J.** (1941) 'Die Logik und das Grundlagenproblem', *Les entretiens de Zürich sur les fondements et la méthode de sciences mathématiques*, 6-9 décembre 1938, 82-100; trans. 'Logic and the Problem of the Foundations of Mathematics', in *Selected Works*, ed. L. Borkowski, Amsterdam: North Holland, 1970. (Referred to in §2.)
- Łukasiewicz, J.** (1951) *Aristotle's Syllogistic from the Standpoint of Modern Formal Logic*, Oxford: Oxford University Press; 2nd enlarged edn, 1957.(Referred to in §2.)
- Łukasiewicz, J.** (1970) *Selected Works*, ed. L. Borkowski, Amsterdam: North Holland.(Anthology of all important papers from the entire creative work of Łukasiewicz.)
- Łukasiewicz, J. and Tarski, A.** (1930) 'Untersuchungen über den Aussagenkalkül', *Comptes Rendus des Séances de la Société des Sciences et des Lettres de Varsovie (Classe 3)* 23: 30-50; trans. 'Investigations into the Sentential Calculus', in J. Łukasiewicz, *Selected Works*, ed. L. Borkowski, Amsterdam: North Holland, 1970; and in A. Tarski, *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*, trans. and ed. J.H. Woodger, Oxford: Clarendon Press, 1956.(Referred to in §2.)
- Luschei, E.C.** (1962) *The Logical Systems of Leśniewski*, Amsterdam: North Holland.(The only book which describes the whole of Leśniewski's scientific achievements; includes a large chapter on terminological explanations and directives. An invaluable and unique guide for those who want to read the originals, but it is not a systematic exposition of Leśniewski's systems. Referred to in §3.)
- Marek, W.** (1977) 'The Foundations of Mathematics in Poland after World War II', in R.O. Gandy and J.M.E. Hyland (eds) *Logic Colloquium '76*, Amsterdam: North Holland, 129-38.(Referred to in §6.)
- Mates, B.** (1953) *Stoic Logic*, Berkeley and Los Angeles, CA: University of California Press.(Referred to in §2.)
- McCall, S.** (ed.) (1967) *Polish Logic 1920-1939*, Oxford: Clarendon Press.(Classic anthology of papers by Ajdukiewicz, Chwistek, Jaśkowski, Leśniewski, Łukasiewicz, Śłupecki and Wajsberg, with notes by Kotarbiński on the development of formal logic in Poland in the years 1900-39; includes part of Jordan (1945).)
- Mostowski, A.** (1952) *Sentences Undecidable in Formalized Arithmetic: An Exposition of the Theory of Kurt Gödel*, Amsterdam: North Holland.(Referred to in §6.)
- Mostowski, A.** (1965) 'Thirty Years of Foundational Studies: Lectures on the Development of Mathematical Logic and the Study of the Foundations of Mathematics in 1930-1964', *Acta Philosophica Fennica* 17: 1-180; repr. New York: Barnes & Noble, 1966; and in *Foundational Studies*, vol. 1, Amsterdam: North Holland, 1979. (History of mathematical logic from Gödel to P.J. Cohen, written from a Polish point of view; referred to in §6.)
- Mostowski, A.** (1979) *Foundational Studies*, Amsterdam: North Holland, 2 vols.(An anthology of the most important of Mostowski's works from 1937-76; includes a complete bibliography and five essays summarizing his work on decidability, recursion and hierarchy theory, set theory, model theory, logical calculi and arithmetic. Volume 1 may be of interest for its survey and expository papers. Referred to in §§5-6.)

- Mostowski, A. et al.** (1955) 'The Present State of Investigations on the Foundations of Mathematics', *Rozprawy Matematyczne (Dissertationes Mathematicae)* 9: 1-48.(Referred to in §6.)
- Pearce, D. and Woleński, J.** (eds) (1988) *Logischer Rationalismus: Schriften der Lemberg-Warschauer Schule (Logical Rationalism: Writings of the Lwów-Warsaw School)*, Frankfurt: Athenaeum.(Shorter papers, in German, by Ajdukiewicz, J. Hosiasson-Lindenbaum, M. Kokoszyńska, Kotarbiński, Łukasiewicz and Tarski.)
- Pepis, J.** (1937) *O zagadnieniu rozstrzygalności w zakresie węższego rachunku funkcyjnego (On the Decision Problem in the Domain of the Restricted Functional Calculus)*, with a German summary, Lwów: Archiwum Towarzystwa Naukowego we Lwowie, §3, vol. 7, fasc. 8, 1-172 ([465]-[616]).(Referred to in §5.)
- Pepis, J.** (1938) 'Untersuchungen über das Entscheidungsproblem der mathematischen Logik' (Investigations into the Decision Problem of Mathematical Logic), *Fundamenta Mathematicae* 30: 257-348.(Referred to in §5.)
- Prior, A.** (1955) *Formal Logic*, Oxford: Clarendon Press; 2nd edn with corrections, 1962.(An example of the use of Polish notation in a textbook.)
- Rasiowa, H. and Sikorski, R.** (1963) *The Mathematics of Metamathematics*, Warsaw: PWN.(Referred to in §6.)
- Salamucha, J.** (1934) 'Dowód "ex motu" na istnienie Boga: analiza logiczna argumentacji św. Tomasza z Akwinu', *Collectanea Theologica* 15: 53-92; trans. 'The Proof *ex motu* for the Existence of God: Logical Analysis of St. Thomas' Arguments', *The New Scholasticism* 32: 334-72, 1958. (Referred to in §2.)
- Salamucha, J.** (1935) 'Logika zdań Wilhelma Ockhama' (Propositional Logic in William of Ockham), *Przegląd Filozoficzny* 38 (3): 208-39.(Referred to in §2.)
- Salamucha, J.** (1937) 'Pojawienie się zagadnień antynomialnych na gruncie logiki średniowiecznej' (The Appearance of the Problem of Antinomies in Medieval Logic), *Przegląd Filozoficzny* 40 (1): 68-89, 40 (3): 320-43.(Referred to in §2.)
- Scholz, H.** (1931) *Geschichte der Logik*, Berlin; repr. as *Abriß der Geschichte der Logik*, Freiburg: Alber, 1959; trans. K.F. Leidecker, *Concise History of Logic*, New York: Philosophical Library, 1961.(Referred to in §2.)
- Sierpiński, W.** (1918) 'L' axiome de M. Zermelo et son rôle dans théorie des ensembles et l'analyse' (Zermelo's Axiom and its Role in Set Theory and Analysis), *Bulletin de l'Académie des Sciences de Cracovie, Classe des Sciences Mathématiques*, série A, 97-152; repr. in *Oeuvres Choisies*, vol. 2, *Théorie des Ensembles et ses Applications*, *Travaux des Années 1908-1929*, Warsaw: PWN-Editions Scientifiques de Pologne, 1975. (Referred to in §5.)
- Sierpiński, W.** (1934) *Hypothèse du Continu (The Continuum Hypothesis)*, Monografie Matematyczne, vol. 4, Warsaw: Grasiński.(Referred to in §5.)
- Sierpiński, W.** (1976) *Oeuvres Choisies (Selected Works)*, vol. 3, *Théorie des Ensembles et ses Applications*, *Travaux des Années 1930-1966 (Set Theory and its Applications: Papers from 1930 to 1966)*, Warsaw: PWN-Editions Scientifiques de Pologne.(Referred to in §5.)
- Sleszyński, J.** (1925-9) *Teoria dowodu. Podług wykładów uniwersyteckich Prof. Dr Jana Sleszyńskiego opracował S.K. Zaremba (The Theory of Proof. Edited by S.K. Zaremba from the University Lectures of Professor Dr. Jan Sleszyński)*, Cracow: Nakładem Kółka Matematyczno-Fizycznego Uczniów Uniwersytetu Jagiellońskiego, 2 vols.(Referred to in §5.)
- Slupecki, J.** (1939a) 'Dowód aksjomatyzowalności pełnych systemów wielowartościowych rachunku zdań', *Comptes Rendus des Séances de la Société des Sciences et des Lettres de Varsovie (Classe 3)* 32: 110-28; trans. 'A Proof of the Axiomatizability of Full Systems of Many-Valued Sentential Calculus', *Studia Logica* 29: 155-68, 1971.(Referred to in §2.)
- Slupecki, J.** (1939b) 'Kryterium pełności wielowartościowych systemów logiki zdań', *Comptes Rendus des Séances de la Société des Sciences et des Lettres de Varsovie (Classe 3)* 32: 102-9; trans. 'A Criterion of Fullness for Many-Valued Systems of Propositional Logic', *Studia Logica* 30: 153-7, 1972. (Referred to in §2.)
- Slupecki, J.** (1949) 'On Aristotelian Syllogistic', *Studia Philosophica (Poznań)* 4: 275-300.(Referred to in §2.)
- Sobociński, B.** (1934) 'O kolejnych uproszczeniach aksjomatyki "ontologii" prof. S. Leśniewskiego', in *Fragmety Filozoficzne. Księga Pamiątkowa ku Uczczeniu Piętnastolecia Pracy Nauczycielskiej w Uniwersytecie Warszawskim Profesora Tadeusza Kotarbińskiego*, Warsaw: Nakładem Uczniów, 144-60; trans. 'Successive Simplifications of the Axiom-System of Leśniewski's Ontology', in S. McCall (ed.) *Polish Logic 1920-1939*, Oxford: Clarendon Press, 1967.(Referred to in §3.)
- Sobociński, B.** (1936) 'Aksjomatyzacja pewnych wielowartościowych systemów teorii dedukcji' (Axiomatization of Certain Many-Valued Systems of the Theory of Deduction), *Roczniki Prac Naukowych Zrzeszenia Asystentów Uniwersytetu Józefa Piłsudskiego w Warszawie* 1: 399-419. (Referred to in §2.)
- Szrednicki, J.T.J., Rickey, V.F. and Czelakowski, J.** (eds) (1984) *Leśniewski's Systems: Ontology and Mereology*, The Hague: Martinus Nijhoff, and Wrocław: Ossolineum.(Twelve representative papers on ontology and mereology by Leśniewski's pupils, Z. Kruszewski, C. Lejewski, J. Slupecki, B. Sobociński; and

- their pupils, J. Canty, R. Clay and B. Iwanuś. Referred to in §3.)
- Surma, S.J.** (1971) 'Jaśkowski's Matrix Criterion for the Intuitionistic Propositional Calculus', *Universitas Iagellonica Acta Scientiarum Litterarumque*, vol. 265, *Schedae Logicae* 6: 21-54. (See Jaśkowski 1936.)
- Surma, S.J.** (1977) 'On the Work and Influence of Stanisław Leśniewski', in R.O. Gandy and J.M.E. Hyland (eds) *Logic Colloquium '76*, Amsterdam: North Holland, 191-220. (A biography and annotated bibliography of Leśniewski, with an outline of his logical systems.)
- Surma, S.J.** (1979) 'On the Origin and Subsequent Application of the Concept of Lindenbaum Algebra', in L.J. Cohen, J. Łoś, H. Pfeiffer and K.-H. Podewski (eds) *Logic, Methodology and Philosophy of Science*, Amsterdam: North Holland, vol. 6, 719-34. (Rather technical exposition of the subject described in the title, with a full bibliography of Lindenbaum's works.)
- Surma, S.J.** (1987) 'The Logical Work of Mordchaj Wajsberg', in J. Szrednicki (ed.) *Initiatives in Logic*, Dordrecht: Martinus Nijhoff, 101-15. (An account of Wajsberg's life and his logical research.)
- Szumilewicz-Lachman, I. and Cohen, R.S.** (ed.) (1994) *Zygmunt Zawirski: His Life and Work. With Selected Writings on Time, Logic and Methodology of Science*, trans. F. Lachman, Dordrecht: Kluwer. (Valuable especially for Zawirski's two papers, 'The Genesis and Development of Intuitionistic Logic' and 'The Evolution of the Notion of Time'; also includes essays by Szumilewicz-Lachman on the Lwów-Warsaw school and the Vienna Circle.)
- Tarski, A.** (1956) *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*, trans. and ed. J.H. Woodger, Oxford: Clarendon Press; repr. and ed. J. Corcoran, Indianapolis, IN: Hackett Publishing Company, 2nd edn, 1983. (Referred to in §§2, 4. For a careful presentation of the content of the book see the review by W.A. Pogorzelski and S.J. Surma, *Journal of Symbolic Logic* 34 (1): 99-106, 1969.)
- Tarski, A.** (1986) *The Collected Papers*, ed. S.R. Givant and R.N. McKenzie, Basle: Birkhäuser, 4 vols. (Reprints of all of Tarski's writings excluding his books; unfortunately original pagination has not been kept.)
- Tarski, A.** (1992) 'Alfred Tarski: Drei Briefe an Otto Neurath', ed. R. Haller, trans. J. Tarski, *Grazer Philosophische Studien* 43: 1-32. (Referred to in §4.)
- Wajsberg, M.** (1933) 'Ein erweiterter Klassenkalkül', *Monatshefte für Mathematik und Physik* 40: 113-26; trans. 'An Extended Class Calculus', in *Logical Works*, ed. S.J. Surma, Wrocław: Ossolineum, 1977. (Referred to in §2.)
- Wajsberg, M.** (1935) 'Beiträge zum Metaaussagenkalkül I', *Monatshefte für Mathematik und Physik* 42: 221-42; trans. 'Contributions to the Meta-Calculus of Propositions', in *Logical Works*, ed. S.J. Surma, Wrocław: Ossolineum, 1977. (Referred to in §2.)
- Wajsberg, M.** (1938) 'Untersuchungen über den Aussagenkalkül von A. Heyting', *Wiadomości Matematyczne* 46: 45-101; trans. 'On A. Heyting's Propositional Calculus', in *Logical Works*, ed. S.J. Surma, Wrocław: Ossolineum, 1977. (Referred to in §2.)
- Wójcicki, R.** (1987) *Theory of Sentential Calculi. An Introduction*, Dordrecht: Reidel. (Referred to in §6.)
- Wójcicki, R.** (1997) 'The Postwar Panorama of Logic in Poland', in M.L. Dalla Chiara *et al.* (eds) *Logic and Scientific Methods*, Dordrecht: Kluwer, 497-508. (Referred to in §6.)
- Woleński, J.** (1989) *Logic and Philosophy in the Lvov-Warsaw School*, Dordrecht: Kluwer. (An authoritative and comprehensive treatise with an excellent bibliography.)
- Woodger, J.H.** (1937) *The Axiomatic Method in Biology*, with appendix by A. Tarski, Cambridge: Cambridge University Press, and New York: Macmillan. (Referred to in §3.)

Predicate Calculus

The central feature of the 'predicate calculus' is a formal language designed to represent arguments involving generalizations. Aristotle had developed a formal logic of 'every', 'some' and 'no', but did not tackle the problem of combining these quantifiers (as they are now called) with relations, or with complex predicates, as in 'Everyone is F or G '. The Stoics had the beginnings of a sentential logic - the logic of the connectives 'and', 'or', 'if' and 'not' - which could later be adapted to deal also with complex predicates. The solution is not, however, a simple matter of pasting these two systems together. It involved four major steps. One was Boole's groundwork for a mathematical system of sentential logic by analogy with an algebra of truth and falsity. Another was De Morgan's move away from the traditional subject-predicate paradigm towards a relational one, so that relations of the form '___ loves ___' could be treated naturally alongside predicates such as '___ is mortal'. The third was the invention by Peirce and Frege of the quantifier-and-variable notation. This solved the twin difficulties of adapting sentential logic to express complex predicates, as in $(\forall x)(Fx \vee Gx)$, and linking the quantifiers unambiguously to the right clauses in multiply general sentences, for example, ' $(\forall x)(\exists y)x$ loves y ' and ' $(\exists y)(\forall x)x$ loves y ' as alternative renderings of 'Everyone loves someone'. Finally there was Frege's break with the traditional semantics of quantified terms. Instead of treating 'every man' as the subject of 'Every man is mortal', he treats it as a second-level predicate governing the first-level predicate 'is mortal'. At one stroke this obviates the futile search for an entity answering to 'everyone' and provides the rationale for disambiguating multiply general sentences.

Frege and Peirce catered for quantification not merely over individuals but over properties, relations and functions of individuals, and Whitehead and Russell carried this further to cover quantification over relations of relations, and so on. These would now be called second-order or higher-order predicate calculi as opposed to the first-order predicate calculus - often simply called 'the' predicate calculus - with which this entry is primarily concerned.

One can give a mathematically precise characterization of the valid argument forms of the predicate calculus. That has become a kind of standard for demonstrating the logical validity of informal deductive arguments, and the semantics of the predicate calculus is the starting point for contemporary philosophy of language. Moreover, the analysis of argument that leads to and incorporates the predicate calculus is the most successful piece of philosophical analysis completed in modern times, and as such serves as a paradigm - for some to be emulated, for others to be avoided, but in one way or another central to the development of contemporary philosophy.

This entry concentrates on the central theme of the analysis of logical validity through an account of what it is for an argument 'form' of the predicate calculus to be valid.

1 Language of the predicate calculus

The notation used here is that of the Hilbert school. A commonly used alternative, derived from Peano and Russell, employs $\sim$ for negation, $\&$ for conjunction, $\supset$ for the conditional, $\equiv$ for the biconditional and $(\forall)$ for the universal quantifier.

We build up the formulas of the predicate calculus from atomic formulas using the sentential connectives and quantifiers. To start with we need a stock of enough variables: $v_0, v_1, \dots, v_n, \dots$. The language may contain constant symbols and relation symbols. (One can also allow function symbols, which would be necessary, for example, for a standard treatment of arithmetic.) Let the constant symbols be $c_0, c_1, \dots, c_n, \dots$. Although the sentential connectives and quantifiers are often referred to as logical constants, in this entry, 'constant' always means a non-logical symbol serving as a name or potential name for an individual. Let the relation symbols be $R_0^n, R_1^n, \dots, R_m^n, \dots$. Those with a superscript of n are called n -placed and will be used for n -ary relations. One-placed or 'monadic' relation symbols may be called predicate symbols. It is convenient to bring the sentence letters of the sentential calculus (see §3) under this uniform notation by regarding them as zero-placed relation symbols. A 'term' is either a variable or a constant. (It will be assumed here that the language of the predicate calculus contains, as well as any non-logical relation symbols, the equality sign '=' used to denote the logical relation of identity. Some writers would call this 'predicate calculus with identity', keeping 'predicate calculus' for the version in which the equality sign is absent or in which identity is not given the special status of a logical

relation.)

An ‘atomic formula’ is an expression either of the form $s = t$, where s and t are terms, or of the form $R^n v_1 v_2 \dots v_n$, where $v_1, \dots, v_n$ are terms, for some n . For example, $R^2 vc$ might be used for ‘ v is the daughter of Margaret Sanger’, where R^2 stands for the daughter relation and c stands for Margaret Sanger. (Here and elsewhere we shall omit the official apparatus of subscripts for the sake of simplicity of exposition.)

The constants and relation symbols function here as schematic letters that may receive varying interpretations, but there is nothing to block fixing the interpretations of some of them in a definite way, which results in an interpreted language of a sort more closely related to ordinary language. The system with only schematic letters, those without a fixed interpretation, is known as the ‘pure’ predicate calculus, while systems with symbols that have fixed interpretations are known as ‘applied’ predicate calculi. Frege considered only applied predicate calculi without any schematic letters. Hilbert introduced and emphasized the importance of the device of schematic letters.

If ϕ and ψ are formulas, then so are those expressions formed as indicated using the sentential connectives: $\neg\phi$, $\phi \wedge \psi$, $\phi \vee \psi$, $\phi \rightarrow \psi$ and $\phi \leftrightarrow \psi$. If ϕ is a formula and x is a variable, then $(\forall x)\phi$ and $(\exists x)\phi$ are *quantified* formulas and the occurrence of ϕ is said to be the ‘scope’ of the quantifier. The symbols are formal counterparts of the English ‘not’, ‘and’, inclusive ‘or’, ‘if... then’, ‘if and only if’, ‘for all’, and ‘there exists at least one’, respectively. An occurrence of a variable x in a formula that is not in the scope of any quantifier over x (that is, a quantifier of the form $(\forall x)$ or $(\exists x)$) in the formula is said to be a ‘free’ occurrence, and a quantifier over the variable x in a formula is said to ‘bind’ every occurrence of x free in its scope. Such an occurrence (including the occurrence of x inside the quantifier itself) is then said to be ‘bound’ by the quantifier.

Some formulas are incomplete expressions, such as $R^2 vc$, which might be used for ‘ v is a daughter of Margaret Sanger’, in which v is left free, while others are complete sentences and may be true or false on their own, such as $(\exists v)R^2 vc$, ‘Margaret Sanger has a daughter’ (literally, ‘There exists v such that v is a daughter of Margaret Sanger’). That distinction can be made precise as follows: a variable ‘occurs free’ in a formula ϕ just if any of its occurrences in the formula is free. A ‘sentence’ is a formula in which no variable occurs free. Thus, while the sentential connectives may on occasion indeed connect sentences, they may also connect formulas that are not sentences.

2 Logically valid arguments and interpretations

An argument is ‘valid’ if it is impossible that the premises be true and the conclusion false. Validity is the central feature of good deductive arguments, for if an argument is valid and the premises true, it follows that the conclusion must also be true. Thus, valid arguments are exactly those suitable for what may be the paradigm application of deductive arguments: establishing the truth of the conclusion on the basis of the truth of the premises. The conclusion of a valid argument is said to be a ‘tautological consequence’ of the premises.

There are other desirable features for an argument to have: to establish the truth of its conclusion, the premises must be true; and to be part of an explanation of its conclusion, its premises must in some sense be better known or understood than the conclusion, and, in particular, the argument must not be circular. We do not discuss such secondary features further here. For the most part, they are only requisites of arguments employed for specific purposes, unlike validity, which is a condition for acceptability of any deductive argument. The notion of validity relies on a notion of possibility that is surely in need of further analysis (but see below).

There are arguments whose validity is not purely logical. For example, ‘Caila is a female child, therefore Caila is a girl’ is valid because a female child is necessarily a girl, although that fact does not appear in the argument. This argument has the ‘form’ ‘ $F(c) \wedge C(c)$, therefore $G(c)$ ’. But so does ‘Caligula was a famous caesar, therefore Caligula was a god’, which is clearly not valid, since the premise is true and the conclusion false. In contrast, the argument ‘Everyone is mortal, therefore Socrates is mortal’ is itself valid and so is every other argument of the form ‘ $(\forall x)M(x)$, therefore $M(s)$ ’. Here the validity of the argument does not depend on anything beyond its form, and in particular not on its subject matter: the reasons for its validity are ‘topic neutral’. That has often been taken to be a criterion for something to be logical. (See Haack (1978) for a historical summary of views on what counts as logical, and Sher (1991) and Koslow (1992) for recent work on the subject.)

The primary business of logic is to characterize those arguments which are not only valid, but of a form such that

every argument of that form is also valid; briefly, those which are ‘logically valid’. It is obvious that logical validity depends on a prior notion of form (see **Logical form**). The sentences of the pure predicate calculus are designed to serve as representatives of forms of sentences of natural language, and the notion that an argument is of a certain form can be made more precise using the idea of a ‘natural-language interpretation’. For example, ‘Caila is a female child, therefore Caila is a girl’ is of the form ‘ $F(c) \wedge C(c)$, therefore $G(c)$ ’ under the interpretation that interprets ‘ c ’ as ‘Caila’, ‘ F ’ as ‘is female’, ‘ C ’ as ‘is a child’ and ‘ G ’ as ‘is a girl’; but it is also of the form ‘ A , therefore B ’ under the interpretation that interprets ‘ A ’ as ‘Caila is a female child’ and ‘ B ’ as ‘Caila is a girl’.

The same argument may have many forms. In general, a natural-language interpretation of an argument form must specify a name or definite description for each constant and a predicate for each predicate symbol. A predicate amounts to little more than a sentence with a single blank in it: ‘__ is female’. Each n -placed relation symbol will similarly be interpreted using an n -ary ‘sentence form’ - a sentence with n blanks in it. A binary relation symbol might be interpreted using, for example, ‘__ is spicier than __’. One needs some mechanism for specifying which term following a relation symbol corresponds to which blank in the sentence form that is its interpretation, but such details will be omitted here.

It follows from the stipulations above that an argument is logically valid if and only if it is of a form that yields a valid argument under all natural-language interpretations. Such a form will be called a valid argument form, and the problem of determining which arguments are logically valid thus amounts to that of determining which argument *forms* are valid.

It is tempting to suppose that an argument form is valid just in case it has no natural-language interpretation that yields an argument with true premises and a false conclusion. That would eliminate the troublesome notion of possibility from the basis of logic. Indeed, Carnap (1934) made proposals along closely related lines. Unfortunately, such a proposal does not serve as an adequate criterion for validity: as Tarski (1936) pointed out, it might be that our language happened not to be rich enough to give rise to the requisite variety of interpretations, and thus the validity of an argument could depend on the richness of parts of the language that were not used in the argument in question. It can in fact be shown that the language of arithmetic is rich enough to yield all the interpretations one could need for the predicate calculus presented here (see Kleene 1952 and Hasenjäger 1953), but it would be question-begging to assume it at the present stage.

3 Sentential logic

Since validity is defined in terms of whether or not the premises of an argument form can be true while the conclusion is not, it is first necessary to specify under what circumstances a sentence is true. We assume every sentence has a ‘truth-value’ of either T or F, where F is anything other than true - which may include, depending on one’s view, not only false, but also meaningless, undefined and perhaps other possibilities. (Any sentence that is not true will be referred to, briefly though perhaps inaccurately, as false. For calculi allowing more than these two truth-values, see **Many-valued logics**.)

It is illuminating to begin with the simple case of ‘sentential logic’ (also called the propositional calculus); that fragment of the predicate calculus without variables, constants or quantifiers, and in which one only allows zero-placed relation symbols. According to the definitions given in §2, an atomic sentence is formed by such a symbol followed by zero terms. It thus amounts to no more than a single ‘sentence letter’. It is possible to specify the form of some arguments in sentential logic. For example, ‘Jane Austen and Charlotte Brontë were novelists, therefore Jane Austen was a novelist’ has the form ‘ $A \wedge B$, therefore A ’. But many arguments cannot be formalized in sentential logic so as to be valid. For example, ‘Everyone is mortal, therefore Hannah Arendt is mortal’, has the form ‘ $(\forall v)R^1(v)$, therefore $R^1(c)$ ’, but this is not a form available in sentential logic. The only form available in sentential logic is the invalid ‘ A , therefore B ’.

The truth-value of a sentence of sentential logic is determined by an assignment of truth-values to its constituent sentence letters, through the following inductive rules. The negation of a sentence is false if the sentence is true, and true if the sentence is false. The conjunction of two sentences is true if both are true, and false otherwise. The disjunction of two sentences is true if at least one is true, and false otherwise. A conditional $\phi \rightarrow \psi$ is true unless ϕ is true and ψ is false. A biconditional between two sentences is true if both sentences have the same truth-value

and false otherwise. The idea of treating the truth-value of complex sentences as a function of the truth-values of their components (see **Compositionality**) - the idea of ‘truth-functional’ connectives - dates back to the Stoics (see **Logic, ancient §5**). The treatment of negation, conjunction and disjunction as truth-functional is scarcely controversial, but the same cannot be said of the truth-functional treatment of the conditional and biconditional (see **Counterfactual conditionals**; **Indicative conditionals**).

The rules just given can be represented diagrammatically in the following ‘truth table’:

ϕ	ψ	$\neg\phi$	$\phi \wedge \psi$	$\phi \vee \psi$	$\phi \rightarrow \psi$	$\phi \leftrightarrow \psi$
T	T	F	T	T	T	T
T	F	F	F	T	F	F
F	T	T	F	T	T	F
F	F	T	F	F	T	T

Let us use the table to compute the truth-value of $((A \leftrightarrow B) \wedge A) \rightarrow (\neg B)$, given values for A and B of T and F, respectively. The truth-value of $A \leftrightarrow B$, according to the table, is F. But then, applying the rule for $\wedge$ to that and the given value for A , the truth-value for $(A \leftrightarrow B) \wedge A$ is also F. Since the truth-value of B is F, that of $\neg B$ is T. Finally, applying the rule for $\rightarrow$ to the last two facts, the truth-value of the whole sentence is T. This computation is represented by the following line of a truth table.

A	B	$A \leftrightarrow B$	$\neg B$	$(A \leftrightarrow B) \wedge A$	$((A \leftrightarrow B) \wedge A) \rightarrow (\neg B)$
T	F	F	T	F	T

A sentence that is assigned the value T on every line of its truth table is known as a ‘tautology’, a term introduced by Wittgenstein, who emphasized the importance of the method of truth tables.

Suppose Γ is a finite set of sentences and ϕ a sentence of sentential logic. It is then possible to construct a truth table that lists the truth-values of ϕ and all the sentences in Γ under all possible combinations of truth assignments to the sentence letters they contain. Now, suppose there is no line of that truth table in which all the sentences in Γ receive the truth-value T and the sentence ϕ receives the value F. Then the corresponding argument form ‘ Γ , therefore ϕ ’ is valid. For consider an arbitrary natural-language interpretation. The argument it yields must be valid - if not it would have to be possible for the premises to be true and the conclusion false. But that is not possible, by hypothesis. On the other hand, suppose that there is a line of the table on which all the sentences in Γ are true and ϕ is false. Then the argument form is not valid. For choose a line of the table of the type just assumed to exist. Consider a natural-language interpretation of the argument form that interprets each sentence letter by ‘Snow is black’ if it is assigned F on the selected line of the table, and by ‘Snow is white’ if it is assigned T on the selected line. The argument yielded by the interpretation just specified is one in which the premises are actually true and the conclusion is actually false. It is therefore not valid, and so the argument form is not valid.

This discussion demonstrates a remarkable fact: truth tables provide a precise criterion for whether or not an argument form in sentential logic is valid - it is valid just if there is no line of the table on which all the premises have value T and the conclusion has value F. (The troublesome notion of possibility has dropped out of the picture, since all that turned out to be relevant about possibility was the combinatorial list of possible ways for various simple sentences to be true or false.) Truth tables in fact constitute a mechanical decision procedure for whether an argument form of sentential logic is valid. For one can generate the table mechanically and then check it line by line.

Though truth tables provide a decision procedure in principle, it soon ceases to be a practically feasible one, since the number of lines of a truth table doubles with the addition of each sentence letter. The question of whether there is a more efficient procedure, in the sense that the difficulty of the computation grows as, for example, a polynomial function of the number of letters rather than an exponential one, is a version of perhaps the most important question in the theory of computational complexity (see Garey and Johnson 1979).

4 Expressive completeness

It might seem that the logic developed here is unduly limited in that it allows only five connectives. What about, for example, exclusive ‘or’, or the majority-of-three connective that holds among three sentences if and only if at

least two of them are true? The ‘expressive completeness theorem’ for sentential logic shows that the limitation is only an apparent one: every truth-functional connective, specified by any truth table whatever, is definable using those already introduced.

The proof of ‘expressive completeness’ (often called functional completeness) can be illustrated using the majority-of-three connective as an example. The sentence formed by applying it to A , B and C has the same truth table as

$$(A \wedge B \wedge \neg C) \vee (A \wedge \neg B \wedge C) \\ \vee (\neg A \wedge B \wedge C) \vee (A \wedge B \wedge C),$$

a disjunction whose components correspond to the four possible ways in which the majority-of-three connective creates a true sentence, these being read off the lines of the truth table. This example is in ‘disjunctive normal form’, that is, a disjunction of conjunctions of sentence letters and their negations. One can alternatively express an arbitrary truth table by a sentence in ‘conjunctive normal form’, that is, a conjunction of disjunctions of sentence letters and their negations. (The majority-of-three connective becomes $(A \vee B) \wedge (B \vee C) \wedge (C \vee A)$.) Any truth table can similarly be expressed using only ‘and’, ‘or’ and ‘not’.

It follows that any set of connectives that can be used to define ‘and’, ‘or’ and ‘not’ is expressively complete. It is an elementary exercise to show in this way that negation plus any one of $\vee$, $\wedge$ or $\rightarrow$ is expressively complete. The binary connective that forms a true sentence if and only if both components are false (‘not or’) is also expressively complete, as is the one that forms a false sentence if and only if both components are true (‘not and’). These last facts were first published by Sheffer (1913), and expressively complete sets of connectives were classified by Post (1941) and Wernick (1942).

5 Truth and logical consequence

The next task is to extend the account of the truth-conditions of sentences to the full predicate calculus. The truth-value of an atomic sentence is determined by whether the objects denoted by the constants stand in the relation denoted by the relation symbol. This can be expressed mathematically by means of an ‘interpretation function’ that associates an object with each relevant constant symbol and a set of ordered n -tuples of objects with each relevant n -placed relation symbol. (Though natural-language interpretations and interpretation functions play similar roles - they fix how the constant and relation symbols are interpreted - they are not the same: a natural-language interpretation takes each symbol to a suitable item of natural language, while an interpretation function takes each symbol to a suitable set-theoretic object.) For example, the relation on persons specified by ‘__ is shorter than __’ could be symbolized using R^2 , which would then be taken by the corresponding interpretation function to the set S of all ordered pairs of persons in which the first person is shorter than the second. If c were to be interpreted as G.E. Moore and d as Kareem Jabbar, then $R^2 cd$ would take on the value T precisely because the ordered pair $\langle \text{G.E. Moore, Kareem Jabbar} \rangle$ is in the set S . The present method goes beyond that of natural-language interpretations in an important respect: an interpretation function can take a relation symbol to an arbitrary collection of ordered n -tuples, even though it might not be specifiable in ordinary language.

In an applied predicate calculus, one selects certain constant and relation symbols to stand for certain objects and sets of n -tuples, reserving the rest to function as schematic letters. One way of introducing integer arithmetic, for example, involves fixing symbols for zero, ‘less than’ and so on. One can then add to the axioms of a logical system special nonlogical axioms involving those symbols and even restrict the language to eliminate schematic letters.

The truth-value of a universally quantified sentence, say $(\forall v) R^2 cv$ (which, on the above interpretation, reads ‘G.E. Moore is shorter than everyone’), would be easy to characterize if every relevant object were the interpretation of some constant symbol. It would then be true if all the following sentences were: $R^2 cc_0, R^2 cc_1, R^2 cc_2, \dots$. But there is no reason to suppose that every relevant object has a name. Consider ‘All ravens are black’ - surely not every raven has a name. And certainly, no one language could name all the uncountably many real numbers. Nonetheless, each relevant object could be given a name, and so the following works. First, specify a non-empty set to be the domain of the quantifiers (persons, ravens, real numbers, ...). Second, specify a new constant that has not been employed in any of the sentences under consideration. For our

example, the domain is to be persons and the constant might be d . Then, continuing the example, $(\forall v)R^2cv$ is true if R^2cd is true no matter how the interpretation so far specified is expanded by interpreting the constant d as standing for a member of the domain.

It is now possible to give a precise definition of how the truth of sentences is determined. A 'language' (sometimes called a signature or similarity type) is a set of constant and relation symbols. (A language might better be called a vocabulary, but the terminology is standard.) A 'structure' (or 'interpretation') for a language $\mathcal{L}$ is an ordered pair with first member a non-empty collection - the 'universe' of the structure - and second member a suitable (in a sense defined immediately below) function with domain $\mathcal{L}$ - the 'interpretation function' of the structure. (The language is thus determined by the structure, being the domain of the interpretation function.) The interpretation function is required to map every constant symbol in $\mathcal{L}$ to a member of the universe, every sentence letter to a truth-value, every predicate symbol to a set, and every n -placed relation symbol, for n greater than 1, to a collection of ordered n -tuples of members of the universe.

When symbols with a fixed usage are in the language $\mathcal{L}$, one must require that the relevant objects be in the universe and that the interpretation function map such symbols to the correct objects and relations. We can avoid such complications by considering only languages that contain nothing but schematic letters. Arithmetic, for example, is then considered in terms not of a particular language but of a particular structure, and the axioms of arithmetic, because they will be written using schematic letters, will have interpretations and interpretation functions in addition to the main one.

If A and A' are structures with the same universe such that the language of A' includes the language $\mathcal{L}$ of A and possibly some other symbols, and if the interpretation functions of the two structures agree on $\mathcal{L}$, then A' is said to be an 'expansion' of A . When one structure is an expansion of another it differs only in that it may possibly expand the interpretation function to contain some additional symbols in its domain. Note that if d is some definite constant symbol not in $\mathcal{L}$, then for every member a of the universe of A there is an expansion in which the interpretation of d is a .

The truth-value of a sentence in a structure is defined by the following conditions on all structures A and sentences ϕ of the language of A :

- (1) (a) If ϕ is of the form $c_i = c_j$, then the truth-value of ϕ in A is T if the interpretations of c_i and c_j in A (that is, the values of the interpretation function on c_i and c_j) are the same, and the truth-value is F otherwise.
- (b) If ϕ is a sentence letter, its value in A is the truth-value to which it is mapped by the interpretation function.
- (c) If ϕ is of the form $R^n c_1 \dots c_n$ with $n > 0$, then the truth-value of ϕ in A is T if the n -tuple formed from the interpretations of $c_1, \dots, c_n$ (in that order) is a member of the interpretation of R^n (which, recall, is a set of n -tuples) and it is F otherwise.
- (2) If ϕ is a negation, conjunction, disjunction, conditional or biconditional, then its truth-value is determined from those of its constituents just as in sentential logic.
- (3) If ϕ is of the form $(\forall x)\psi$, $\mathcal{L}$ is the language of A and $\mathcal{L}'$ is the language obtained from $\mathcal{L}$ by adding some definite constant symbol d not in $\mathcal{L}$, then the value of ϕ is T if the truth-value of the sentence $\psi(d)$ obtained from ψ by replacing every free occurrence of x by d is T in every expansion of A to a structure for $\mathcal{L}'$. The truth-value is F otherwise.
- (4) If ϕ is of the form $(\exists x)\psi$, $\mathcal{L}$ is the language of A and $\mathcal{L}'$ is the language obtained from $\mathcal{L}$ by adding some definite constant symbol d not in $\mathcal{L}$, then the value of ϕ is T if there is at least one expansion of A to a structure for $\mathcal{L}'$, in which the truth-value of the sentence $\psi(d)$ obtained from ψ by replacing every free occurrence of x by d is T. It is F otherwise.

It is now possible to define, as Tarski did (1936), logical consequence and logical truth for the predicate calculus in a way parallel to the definitions of tautological consequence and tautology given above. Let the language $\mathcal{L}$ contain only schematic letters. Let Γ be a set of sentences of $\mathcal{L}$ and ϕ a sentence of $\mathcal{L}$. Then ϕ is a 'logical consequence' of Γ if there is no structure for $\mathcal{L}$ in which every sentence in Γ has value T and ϕ has value F; and ϕ is a 'logical truth' if it receives value T in every structure for $\mathcal{L}$. As in the case of sentential logic, one can establish that an argument formalized in the predicate calculus is logically valid just if the formalization of the conclusion is a logical consequence of the formalizations of the premises. In this characterization of logical

validity, the precise mathematical notion of a structure has replaced the troublesome notion of logical possibility.

For technical variants of the semantics presented here, including the idea of extending the language by giving the existing variables the role played here by the added constants, see Tarski (1936) and the textbooks by Mates (1965) and Tennant (1978). The customary semantics excludes the possibility of an empty universe or of constants that do not denote any individual, but both can be accommodated at the cost of some complication (see [Free logics](#)).

It is immediate from the definitions that in any structure, if c and d denote the same individual and ϕ is true, then so is the result of substituting d for any or all occurrences of c in ϕ . Likewise, relation symbols denoting the same set of ordered n -tuples, predicates denoting the same set, and sentences with the same truth-value, can all be substituted for one another without changing the truth-value. In short, the predicate calculus is ‘extensional’ through and through. Some natural-language forms are extensional, others not. Since ‘Fibonacci’ and ‘Leonardo of Pisa’ denote the same man and Fibonacci was a mathematician, it follows that Leonardo was a mathematician; but someone may know that Fibonacci was a mathematician without knowing that Leonardo was, so ‘knows that’ generates forms that are not extensional. The predicate calculus is not adequate to deal with such cases (see [Intensional logics](#)), but even within the extensional sphere, there is no counterpart for the predicate calculus of the expressive completeness theorem for sentential logic. It will in general be true that for a given structure there will be relations on the structure not definable in the language of the structure, structures indistinguishable by any sentence, and more classes of structures for a language than there are sentences of the language.

As for the definability of quantifiers, many are indeed definable with the help of the equality sign, including ‘there are n things such that’ for each finite number n . But the quantifier ‘for finitely many’ cannot be defined in the first-order predicate calculus. It requires the resources of second-order predicate calculus, in which the quantifiers range over relations of objects as well as over individual objects (see [Second-and higher-order logics](#)).

6 Formal deduction and decidability

As an alternative to the semantic approach adopted above, the inferential role of the various logical constants can be specified by codifying a system of rules for deduction and proof. The idea goes back to Aristotle, though the first fully rigorous system is that of Frege (1879), with his insistence that there be absolutely no room for doubt or disagreement whether a putative proof is in accordance with the rules or not. Some have gone further and taken purely syntactic rules for the manipulation of the logical constants to be constitutive of their meanings, and systems of this sort have been used to introduce intuitionistic logic as well as the ‘classical’ logic expounded here (see [Intuitionism](#); [Natural deduction, tableau and sequent systems](#)).

There are many varieties of deductive system (see [Formal languages and systems](#)). The axiomatic variety is conceptually the simplest. A number of logical axioms are laid down, together with rules of inference of the form ‘From $\phi_1, \dots, \phi_n$ infer ψ ’ for suitably related ϕ s and ψ . A common choice is to take the rule of *modus ponens*, ‘From ϕ and $\phi \rightarrow \psi$ infer ψ ’, as the sole rule of inference, supplemented by numerous axioms to offset the absence of other rules. A formal deduction of a conclusion from a set of premises is defined as a finite sequence of sentences which ends with the conclusion and of which every member is either a logical axiom or a premise or else follows immediately from preceding members by a rule of inference. Formal proof is introduced as the limiting case of a deduction from the empty set of premises. A sentence is said to be ‘deducible’ from others in the system if there exists a formal deduction of it from them, and a sentence is a theorem if there exists a formal proof of it. Church (1956) gives a survey of axiomatic systems for sentential logic and for predicate logic. The latter generally invoke an extended deductive apparatus that covers formulas with free variables as well as sentences proper, but Quine (1955) presents an axiomatic system confined strictly to sentences.

Natural deduction systems include a more sophisticated kind of rule, designed to exploit assumptions – propositions assumed for argument’s sake and treated as extra premises for as long as they are wanted. Proof by *reductio ad absurdum* is a good informal illustration of the idea: wishing to prove ϕ , one supposes for argument’s sake that it is false and tries to derive a contradiction, whereupon one discards the assumption and infers that ϕ must be true. The definition of formal deduction now has to keep track of the introduction and subsequent discharge (discarding) of assumptions; see, for example, Lemmon (1965). This complication is, however, offset by there being fewer or even no axioms, and by the comparative naturalness and brevity of

deductions. (For this and some related kinds of deductive system see [Natural deduction, tableau and sequent systems §1.](#))

To be adequate for its purpose, any deductive system must satisfy two conditions. It must be ‘sound’, in the sense that it does not license deductions whose conclusions do not follow logically from their premises. And it must be ‘complete’, that is, strong enough to ensure that every logical consequence of a set of sentences is deducible from them. When a system is both sound and complete, deducibility becomes equivalent to logical consequence and likewise theoremhood is equivalent to logical truth. Of the two conditions, completeness is nearly always much more difficult to establish. The first proof of the completeness of a deductive system for the predicate calculus was given by Gödel (1930; see also [Model theory §4](#)), while Henkin (1949) devised a novel method of proving completeness that has found applications far beyond its original one to the predicate calculus. Although second- and higher-order predicate calculi have significantly more expressive capacity than first-order, their great drawback is that there can never be a sound and complete deductive system for them (see [Second- and higher-order logics §1](#)). The dominant position of the first-order calculus is the result of its being rich enough to formalize a vast range of arguments from mathematics and elsewhere, yet simple enough for all its valid argument forms to be reducible to a few simple rules of inference.

If a sentence of the predicate calculus is a logical consequence of others, that fact can be shown by deriving it from them in a sound and complete deductive system. Although in practice this may require considerable ingenuity, in principle a machine can generate a list of successive finite sequences of sentences on which sooner or later the right one - by hypothesis there is a right one - will turn up. Nonetheless, the method of formal deduction does not provide a decision procedure for logical consequence. The fact that a deduction has not yet turned up means nothing. For there is in general no way of telling whether this is because no such deduction exists - that is, the conclusion in question is not a logical consequence of the premises - or whether it is merely that we have not tried hard or long enough. If we could find a counterexample - a structure in which the putative conclusion is false while the premises are true - that would indeed show that it is not a logical consequence of them. But the parallelism between truth tables and structures breaks down at this point. In sentential logic all one has to do is set out the relevant truth table and inspect its rows. There will, however, always be infinitely many relevant structures, not to mention structures with infinitely many objects. The analogue of setting out and inspecting the rows of a truth table thus involves infinitely many steps and cannot be carried out even in principle.

This strongly suggests the absence of a decision procedure for the predicate calculus, and Church has proved that none exists. However, positive results can be obtained in some restricted cases, the simplest and most important example being that there is a decision procedure for the monadic predicate calculus, that is, the fragment whose language includes predicate symbols but no n -placed relation symbols for $n > 1$. (See [Church’s theorem and the decision problem](#).) For further results of this sort, see Dreben and Goldfarb (1979). Many interesting questions concerning feasibility and quantificational logic arise (see Börger 1989).

See also: [Logical and mathematical terms, glossary of](#)

SHAUGHAN LAVINE

References and further reading

- Börger, E.** (1989) *Computability, Complexity, Logic*, New York: North Holland. (A comparatively accessible treatment of the areas of research on the border between mathematical logic and computational complexity theory.)
- Carnap, R.** (1934) *Logische Syntax der Sprache*, Vienna: Springer; trans. A. Smeaton, *Logical Syntax of Language*, London: Kegan Paul, 1937. (An influential discussion of the relation between ordinary language and the sort of formal language introduced here, the notions of logical truth and consequence and much else. Carnap (1935) discusses many of the same ideas in a less technical way.)
- Carnap, R.** (1935) *Philosophy and Logical Syntax*, London: Kegan Paul, Trench, Trubner & Co. (A semi-popular account of Carnap’s attitude towards philosophy. A clear and useful example of one important sort of impact the predicate calculus has had on philosophers for much of the twentieth century.)
- Carnap, R.** (1942) *Introduction to Semantics*, Cambridge, MA: Harvard University Press. (An influential later discussion of the relation between ordinary language and the sort of formal language introduced here, the notions of logical truth and consequence, and much else.)
- Church, A.** (1956) *Introduction to Mathematical Logic*, Princeton, NJ: Princeton University Press. (An influential,

- exceptionally rigorous introduction to the predicate calculus, with a wealth of historical information.)
- Dreben, B. and Goldfarb, W.D.** (1979) *The Decision Problem: Solvable Classes of Quantificational Formulas*, Reading, MA: Addison-Wesley. (A detailed analysis of which fragments of the predicate calculus are simple enough have a decision procedure.)
- Frege, G.** (1879) *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*, Halle: Nebert; trans. 'Begriffsschrift, a Formula Language, Modelled Upon That of Arithmetic, for Pure Thought', in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967, 1-82. (The classic source for the predicate calculus, in the form of second-order predicate calculus.)
- Garey, M.R. and Johnson, D.S.** (1979) *Computers and Intractability: A Guide to the Theory of $\mathcal{NP}$ -Completeness*, San Francisco, CA: Freeman. (A very readable account of why truth tables are not a practical decision method, how the notion of a practical method can be characterized, and related problems of feasibility.)
- Gödel, K.** (1930) 'Die Vollständigkeit der Axiome des logischen Funktionenkalküls', *Monatshefte für Mathematik und Physik* 37: 349-60; trans. 'The Completeness of the Axioms of the Functional Calculus of Logic', in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967, 582-91. (Cited in §5. The first proof of the completeness of an axiom system for the predicate calculus.)
- Haack, S.** (1978) *Philosophy of Logics*, New York: Cambridge University Press. (A thorough and engaging synoptic account of many of the issues raised in this entry and of alternatives and generalizations of the logic introduced here.)
- Hasenjaeger, G.** (1953) 'Eine Bemerkung zu Henkins Beweis für die Vollständigkeit des Prädikatenkalküls der ersten Stufe', *Journal of Symbolic Logic* 18: 42-8. (Includes one of the first proofs that if a sentence of the predicate calculus has a model, it has one with domain the natural numbers and in which all the predicates are definable in the language of arithmetic.)
- Henkin, L.** (1949) 'The Completeness of the First-Order Functional Calculus', *Journal of Symbolic Logic* 14: 159-66. (Cited in §5.)
- Hilbert, D. and Ackermann, W.** (1928) *Grundzüge der theoretischen Logik*, Berlin: Springer, 2nd edn, 1938; trans. L.M. Hammond, G.G. Leckie and F. Steinhardt, *Principles of Mathematical Logic*, ed. R.E. Luce, New York: Chelsea, 1950. (A classic introduction to the predicate calculus, with a wealth of detail not available in later texts.)
- Kleene, S.C.** (1952) *Introduction to Metamathematics*, Amsterdam: North Holland, and New York: Van Nostrand. (Perhaps the most influential modern text on the mathematics of the predicate calculus.)
- Koslow, A.** (1992) *A Structuralist Theory of Logic*, New York: Cambridge University Press. (An insightful attempt to characterize what is logical in terms of an analysis of the consequence relation.)
- Lemmon, E.J.** (1965) *Beginning Logic*, London: Nelson. (Cited in §5. An exceptionally user-friendly exposition of a natural deduction system for the predicate calculus.)
- Mates, B.** (1965) *Elementary Logic*, New York: Oxford University Press. (Cited in §4.)
- Peirce, C.S.** (1931-58) *Collected Papers*, vol. 2, *Elements of Logic*, vol. 3, *Exact Logic*, ed. C. Hartshorne and P. Weiss, Cambridge, MA: Harvard University Press. (Peirce was one of the inventors of quantifiers, and he coined the term.)
- Post, E.L.** (1941) *The Two-Valued Iterative Systems of Mathematical Logic*, Princeton, NJ: Princeton University Press. (Includes an analysis of the complete sets of connectives for sentential logic.)
- Quine, W.V.** (1940) *Mathematical Logic*, Cambridge, MA: Harvard University Press, 1955. (Cited in §5.)
- Sheffer, H.M.** (1913) 'A Set of Five Independent Postulates for Boolean Algebras, With Application To Logical Constants', *Transactions of the American Mathematical Society* 14: 481-8. (The first publication of the fact that there is a single connective that is expressively complete for sentential logic.)
- Sher, G.** (1991) *The Bounds of Logic: A Generalized Viewpoint*, Cambridge, MA: MIT Press. (A thorough working out of the Tarskian idea that what counts as logical should be characterized in terms of what can be characterized using semantics of the sort described in §4.)
- Tarski, A.** (1933) *Pojęcie prawdy w językach nauk dedukcyjnych*, Warsaw; trans. J.H. Woodger (1956), 'The Concept of Truth in Formalized Languages', in *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*, ed. J. Corcoran, Indianapolis, IN: Hackett Publishing Company, 2nd edn, 1983, 152-278. (One of the most influential works of the century; includes the definition of truth presented here.)
- Tarski, A.** (1936) 'O pojęciu wynikania logicznego', *Przegląd Filozoficzny* 39: 58-68; trans. J.H. Woodger (1956), 'On the Concept of Logical Consequence', in *Logic, Semantics, Metamathematics: Papers from 1923*

to 1938, ed. J. Corcoran, Indianapolis, IN: Hackett Publishing Company, 2nd edn, 1983.(The definitive formulation of the semantic conception of consequence.)

Tennant, N. (1978) *Natural Logic*, Edinburgh: Edinburgh University Press.(Cited in §4.)

Wernick, W. (1942) 'Complete Sets of Logical Functions', *Transactions of the American Mathematical Society* 51: 117-32.(An analysis of the complete sets of connectives for sentential logic.)

Whitehead, A.N. and Russell, B.A.W. (1910) *Principia Mathematica*, vol. 1, Cambridge: Cambridge University Press; 2nd edn, 1925; repr. 1994.(This is the work that established that, in a suitable sense, all of classical mathematics can be formalized within the higher-order predicate calculus.)

Predication

Some sentences have a very simple structure, consisting only of a part which serves to pick out a particular object and a part which says something about the object picked out. Expressions which can be used to say something about objects picked out are called predicates. Thus 'smokes' in 'Sam smokes' is a predicate. But 'predication' may refer either to the activity of predicating or to what is predicated. To understand either we need to know what predicates are and how they combine with other expressions.

Predicates, unlike proper names, can be negated. They combine with other expressions in ways described by categorial grammar: predicates are incomplete and are completed by other expressions, such as proper names. The word 'smokes' in 'Sam smokes' is called a monadic or 1-place predicate; predicates with two or more places are called relational predicates ('Sam loves Erna', '3 is between 2 and 4').

Since Frege it has been customary to hold that the incomplete or predicative parts of 'Sam smokes' and 'Sam is a smoker' are, respectively, 'smokes' and 'is a smoker' (see Frege, G. §§2-4). According to an older view, the incomplete part of 'Sam is a smoker' is the copula 'is', which is completed by two complete expressions, 'Sam' and 'a smoker' ('a' can be ignored as an accident of English). Sometimes called 'the two name theory of predication', this view allows two types of name - proper names and common nouns.

Is predication merely a matter of words? Are mental acts or the senses of words essential to it? For Frege, names and predicates are correlated with ideal senses, and the sentential wholes to which they belong with ideal thoughts or propositions (see [Sense and reference §2](#); Frege 1892a). The sense or meaning of a proper name is sometimes called an individual concept, that of a predicate a concept (Frege himself used neither 'concept' nor 'meaning' in this way). It has been held that to use a predicate is to perform a mental act of predicating and that to use a proper name is to perform a mental act of naming or referring. Husserl and others appeal to both senses and mental acts and conceive of the latter as tokenings of the former. The unity of thoughts and of thinkings, on these views, resembles but is prior to the unity of the sentence: predicative senses require non-predicative senses; predicatings require referrals or acts of quantifying over individuals.

Predication occurs in the context of a variety of mental acts and linguistic actions: I may wonder or ask whether Sam smokes. So to predicate is not always to judge or assert. To assert that if Sam smokes he will smell is not thereby to assert that Sam smokes, although it is to predicate this of him in the context of entertaining or supposing the thought that Sam smokes.

Frege and Husserl pointed out that to deny that Sam smokes is just to assert that he does not smoke. Denying and judging are not on a par because 'not' belongs within the sentence, its sense to that of the sentence. There is, it is true, the phenomenon of polemic negation, as when the third word of the following sentence is stressed: 'Wittgenstein was not German'. But this is a pragmatic phenomenon.

What is it that is predicated? Just as names designate things, so too predicates are said to stand in a semantic relation (reference, signification) to one or more of the following: properties (attributes); what Frege (in his own technical sense) called 'unsaturated concepts'; sets; and relations between objects and possible worlds (see [Semantics, possible worlds §9](#)). Thus predication may be held to involve saying of Sam that he falls under the sense of the predicate 'smokes', exemplifies the property of being a smoker, or belongs to the set of smokers - or some combination of these. The semantic values of predicates and other parts of a sentence are sometimes said to form states of affairs or situations - which may or may not obtain - or facts.

The foregoing contains five simplifications and questionable presuppositions. First, does predication only occur within sentences (thoughts, thinkings)? If a builder shouts 'Slab!' to his assistant is he not producing a non-elliptic predication? Second, we may wonder, with Wittgenstein (1968), to whom the preceding example is due, whether there is in fact any uniform type of semantic relation between predicates, on the one hand, and properties and their ilk, on the other hand.

Third, the apparent (surface, grammatical) forms of predicates can diverge in different ways from their logical (deep) forms. To say that Pierre is tall is not to predicate the monadic property of being tall of him but to predicate,

of him and the average member of some reference class, the relational property of being taller than. Furthermore, if we accept the view (of Davidson; and of Parsons 1990) that many sentences involve quantification over events, then to say that Sam is smoking is to predicate of some event that it is an event of smoking and to predicate of it and Sam that Sam is the agent of it (see [Adverbs §1](#)).

Fourth, according to Ramsey (1990) there is no essential distinction, in ‘Socrates is wise’, between its subject and monadic predicate because the sentence expresses the same proposition as ‘Wisdom is a characteristic of Socrates’. So no fundamental classification between individuals and properties can be based upon such a distinction.

Finally, if we take seriously the claim that predicates are incomplete, then we should say that in ‘Sam smokes’ the predicate is not ‘smokes’ but ‘smokes’, the gappy result of deleting the proper name from the sentence. Thus both ‘Booth shot Lincoln’ and ‘Booth shot Booth’ contain the predicate ‘Booth shot’. But, as Geach indicates in his discussion of this example (‘Quine on Classes and Properties’, 1981: 222-5), the predicate here is not a physical part of the two sentences; it is a sentence pattern which ‘Booth shot’ helps to form. A related view is that such sentence patterns are abstract features of sentences, as melodies are of spoken sentences. But then we might wonder whether ‘Booth’ is not just as much a sentence pattern as ‘Booth shot’.

See also: [Logical and mathematical terms, glossary of](#); [Proper Names](#); [Property theory](#); [Strawson, P.F. §5](#)

KEVIN MULLIGAN

References and further reading

- Frege, G.** (1891) *Funktion und Begriff*, Jena: Pohle; trans. P.T. Geach, ‘Function and Concept’, in *Collected Papers on Mathematics, Logic and Philosophy*, ed. B. McGuinness, Oxford: Blackwell, 1984, 137-56.(Classic account of the role of functions in predication.)
- Frege, G.** (1892a) ‘Über Sinn und Bedeutung’, *Zeitschrift für Philosophie und philosophische Kritik* 100: 25-50; trans. M. Black, ‘On Sense and Meaning’, in *Collected Papers on Mathematics, Logic and Philosophy*, ed. B. McGuinness, Oxford: Blackwell, 1984, 157-77.(The classic account of the sense and reference of sentences and their parts.)
- Frege, G.** (1892b) ‘Über Begriff und Gegenstand’, *Vierteljahrsschrift für wissenschaftliche Philosophie* 16: 192-205; trans. P.T. Geach, ‘On Concept and Object’, in *Collected Papers on Mathematics, Logic and Philosophy*, ed. B. McGuinness, Oxford: Blackwell, 1984, 182-94.(Classic account of what predicates and names signify.)
- Geach, P.T.** (1975) ‘Names and Identity’, in S. Guttenplan (ed.) *Mind and Language*, Oxford: Clarendon Press, 139-58.(Important treatment of the nature of predication, drawing on Frege.)
- Geach, P.T.** (1981) *Logic Matters*, Oxford: Blackwell.(See ‘History of the Corruptions of Logic’ for a historical sketch of the relations between the two name theory and the theory fully worked out by Frege.)
- Gardies, J.-L.** (1975) *Esquisse d’une grammaire pure*, Paris: Vrin; trans. K. Mulligan, *Rational Grammar*, Munich: Philosophia, 1985.(Useful guide to categorial grammar and its applications to natural languages.)
- Parsons, T.** (1990) *Events in the Semantics of English*, Cambridge MA: MIT Press.(Elegantly defends the controversial view that many sentences involve quantification over and predication of events and states.)
- Ramsey, F.** (1990) ‘Universals’, in *Philosophical Papers*, ed. H. Mellor, Cambridge: Cambridge University Press, 8-30.(Stimulating, critical examination of some ways of distinguishing between subjects and predicates and between particulars and properties.)
- Simons, P.** (1981) ‘Unsaturatedness’, *Grazer Philosophische Studien*, 14: 73-95.(Surveys different types of unsaturatedness and incompleteness and their relations to the notions of function, functor and predication.)
- Strawson, P.F.** (1974) *Subject and Predicate in Logic and Grammar*, London: Methuen.(Surveys fully most aspects of predication.)
- Wittgenstein, L.** (1968) *Philosophical Investigations*, Oxford: Blackwell.(Includes discussion of a series of examples that throw doubt on the plausibility of many common assumptions about sentences and semantic relations.)

Presupposition

There are various senses in which one statement may be said to 'presuppose' another, senses which are in permanent danger of being confused. Prominent among them are Strawsonian presupposition, a relation which obtains between statements when the falsity of one deprives the other of truth-value (for example, 'There was such a person as Kepler' is a Strawsonian presupposition of 'Kepler died in misery'); semantic presupposition, which obtains between a statement and a particular use of a sentence type, when the falsity of the statement means that that use will not after all constitute the making of a statement (for example, 'The name "Kepler" has a bearer' is a semantic presupposition of 'Kepler died in misery'); and pragmatic presupposition, a broader notion exemplified by the legitimate presumption that accepting or denying the statement 'Fred knows that the earth moves' means accepting 'The earth moves'.

1 Introduction

In most conversations a certain amount of knowledge is assumed by the parties. Conversing with people whom we know well, we can take it for granted that they know who we are and what we have recently been doing, along with much else. Ordinarily, 'presuppositions' are nothing other than these assumed parts of the conversational background.

The notions of presupposition which have been of greatest concern to linguists and philosophers, however, have been more narrowly defined (but see §4 below), typically being concerned with the way certain phrases or constructions indicate that the speaker is making a presupposition in the ordinary sense. So, if I affirm 'Wellington's victory at Waterloo was his greatest triumph', my form of words indicates that I am presupposing that Wellington was the victor at Waterloo and asserting that it was his greatest triumph. However, the commonly stressed opposition between presupposition and assertion is in some measure misleading. In asking 'Is Tony going to Salzburg again this summer?', I presuppose that he has been there before and inquire whether he intends to go there in the future; no assertion has been made.

This phenomenon may be closely related to another. If I affirm

(1) All John's children are asleep,

I presuppose that John has children and assert that they are asleep. In affirming

(2) If all John's children are asleep, we should not leave the house,

I do not make the same assertion, but I do still presuppose that John has children. It is this that linguists have in mind when they speak of presuppositions being easily 'inherited': they are often (but not always) retained even when the clauses that convey them are sententially embedded so as to lose their own assertive force. Much of the work done by linguists on this topic has been concerned with identifying constructions in which presuppositions do not survive, and to explain why inheritance is blocked (see various authors collected in [Davies 1991](#)).

Philosophers writing on the subject, by contrast, have largely been concerned to elucidate the nature of presupposition: to explain what one thing's presupposing another really amounts to; and to specify the kinds of item between which the relation of presupposition obtains.

2 Strawsonian presupposition

The writer most responsible for bringing the notion of presupposition to the attention of philosophers is P.F. [Strawson](#), from whose work (1) above is drawn. Commenting on the situation of somebody who makes an assertion by uttering (1), Strawson remarks that

he will not normally, or properly, say this, unless he believes that John has children.... But suppose...John has no children. Then is it true or false that all John's children are asleep? Either answer would seem to be misleading. But we are not compelled to give either answer. We can, and normally should, say that, since John has no children, the question does not arise.

(1952: 173-4)

Strawson is eventually led to advance the following general explanation: ‘if [statement] *S* is a necessary condition of the *truth or falsity* of [statement] *S*,...then *S presupposes* *S*’ (1952: 175; original emphasis). So the statement that John has children is a ‘Strawsonian presupposition’ of the statement that all John’s children are asleep; for unless the former is true the latter is neither true nor false. Strawson’s talk of a ‘question [which] does not arise’ is liable to mislead. If John has no children, then the question ‘Is statement (1) true or false?’ may be raised; but the correct answer will be that it is neither.

To understand the notion of Strawsonian presupposition properly, it is necessary to understand what exactly a statement is. Strawson distinguishes between declarative ‘sentences’ - linguistic types the use of which on various occasions of utterance enables people to say things - and the ‘statements’ thereby made. It is statements, not sentences, that are true or false. Should a presupposition fail, there will be a ‘truth-value gap’: a statement will qualify neither as true nor false. An analogous notion applies to other kinds of speech act. So a yes/no question *Q* may be said to presuppose a statement *S* if the truth of *S* is a necessary condition of *Q*’s being answered by ‘yes’ or ‘no’ (see Strawson 1954).

The notion of Strawsonian presupposition has certainly found important philosophical application. In Strawson’s own writings it forms the basis of a counter to Russell’s suggestion that simple sentences involving ‘definite descriptions’ - that is, sentences of the form ‘The *F* is *G*’ - are not of subject-predicate form (see Descriptions). Russell held that such a sentence is best analysed as an existential sentence, ‘There is one and only one *F*, and it is *G*’, whereby the statement that the *F* is *G* straightforwardly entails the existence of an *F*. Having attacked Russell’s arguments in its favour, it was on precisely this score that Strawson tried to refute the theory itself: he insisted that the statement that the present king of France is bald presupposes (rather than entails) that there is a present king of France (see Descriptions §4).

In assessing this suggestion, note that when contemplating statements which are neither true nor false, Strawson requires ‘false’ to bear a sense other than ‘not true’. This means that, in the first place, the difference between Strawson and Russell is less than might at first have been supposed. Since Strawson accepts Russell’s account of the conditions under which a statement made by uttering ‘The *F* is *G*’ will be true, he agrees *ipso facto* with his account of the conditions under which such a statement is not true. They differ only over the conditions for such a statement to be false, Strawson deeming such statements to be neither true nor false in circumstances where Russell would deem them straightforwardly false.

In the second place, however, we are driven to wonder what Strawson means by calling a statement ‘false’, given that it is not a synonym for ‘not true’. It is reasonable to expect the attribution of falsity to a statement to be regulated by the principle that *S* is false just in case the negation of *S* is true, and this is a principle that Strawson’s treatment respects; indeed, it seems that he takes this principle to be explanatory of the notion of falsity. But then it becomes crucial to have a test for identifying the negation of a statement, especially for natural languages, where the syntax does not necessarily make this obvious. Since Strawson’s reluctance to classify ‘The king of France is bald’ as false when there is no king of France stems from a desire to reserve that attribution for the case in which there is a king of France and he is not bald, it is plain that Strawson regards the latter case as the proper negation of this statement. This, indeed, is what one might expect, given that Strawson is defending the view that ‘The *F* is *G*’ is a subject-predicate sentence. But the Russellian will want to draw a semantic distinction between ‘It is not the case that the king of France is bald’ and ‘The king of France is not bald’. This suggests that the issue between them ultimately turns, not upon questions concerning presupposition, but upon the proper treatment of negation (and other logical modifiers) in natural languages.

3 Semantic presupposition

A rather different notion of presupposition is suggested by Frege: ‘If anything is asserted, there is always an obvious presupposition that the simple or compound proper names used have a reference [*Bedeutung*]. If therefore one asserts “Kepler died in misery”, there is a presupposition that the name “Kepler” designates something’ (1892: 40). In order to see the difference between this notion of presupposition and Strawson’s, it helps to make one assumption about statements; namely, that the same statement can be made in a variety of languages. Suppose, then, that there is a sequence of Chinese characters by the inscription of which it may be stated that Kepler died in misery. Such an inscription will certainly not contain the name ‘Kepler’, so it will not impede the making of that

statement, should the English (or German) name 'Kepler' lack a reference. But in that case,

(3) The name 'Kepler' has a reference

plainly is not a Strawsonian presupposition of

(4) Kepler died in misery,

that is, of the statement made by uttering (4). For the latter statement could be made by inscribing the Chinese characters instead, and the statement thereby made could have a truth-value even if (3) were false. What does qualify as a Strawsonian presupposition of (4) is the following statement:

(5) There was such a person as Kepler.

We need, then, to distinguish the new notion of presupposition, exemplified by (4) and (3), from that exemplified by (4) and (5).

In explicating this new notion, it helps to invoke again Strawson's distinction between a (declarative) sentence type - which is the bearer of meaning - and the statement made when the sentence is uttered or used - which is the bearer of truth-value. For with that distinction in mind, we see that the new notion is not a relation between statements, but a relation between a sentence - as used (with a particular meaning) upon a particular occasion - and a statement. Thus, the sentence 'You are drunk' (as used on a particular occasion) presupposes that the speaker is addressing somebody, in the sense that if not, then the use of that sentence on a particular occasion will not constitute the making of a statement. Generally, then:

A use of sentence σ in circumstance c presupposes statement S in so far as the truth of S is necessary for that use to qualify as the making of a statement.

When S and σ are so related, we may say that S is a 'semantic presupposition' of sentence σ , relative to c .

4 Pragmatic presupposition

Whatever the differences between them, Strawsonian and semantic presupposition are alike in depending upon the content of statements, or the meanings of sentence types. That is to say, they are alike in abstracting from the particular conversational purposes of the person uttering the sentence. In this respect they differ from another presupposition relation in which the actual speaker looms much larger. Following Stalnaker (1974), we might say that

A statement S' is a 'pragmatic presupposition' of a statement S in so far as a hearer may reasonably infer that the speaker accepts S' either from their acceptance of S or from their denial of it.

One salient difference between this explication and the definition of a Strawsonian presupposition is that the notions of truth and falsity have been replaced by those of acceptance and denial (acceptance of the negation). Another is the introduction of the notion of a hearer's reasonable inference - reasonable, that is, by the lights of a well-run conversation.

Whether one statement is a pragmatic presupposition of another depends crucially, then, upon what constitutes a well-run conversation, a notion philosophers have tried to explicate by formulating general 'conversational maxims', such as the rule (R) that one's hearer may assume that one is making the logically strongest relevant statement compatible with one's beliefs. To illustrate this, and to show how a pragmatic presupposition might obtain even when a Strawsonian presupposition does not, suppose that speakers are known to abide by (R), and consider the statements S : 'Fred knows that the earth moves' and S' : 'The earth moves'. S' is not a Strawsonian presupposition of S : when S' is false, we may conclude simply that Fred's belief is not in fact knowledge and that S itself is straightforwardly false. All the same, a strong case can be made for the contention that S' is a pragmatic presupposition of S : I may reasonably infer that a speaker accepts S' , should I hear them either accepting or denying S (see Stalnaker 1974). The case of accepting S presents no difficulty; and a denial of S amounts to an assertion of 'Fred doesn't know that the earth moves', an assertion which I may suppose that the speaker would not have made had they accepted the stronger statement 'The earth does not move'.

See also: [Descriptions](#); [Implicature](#); [Logical and mathematical terms, glossary of](#); [Pragmatics](#)

IAN RUMFITT

References and further reading

- Davies, S.** (1991) *Pragmatics: A Reader*, New York: Oxford University Press. (Includes many papers showing the nature of contemporary work on presupposition by linguists, including articles on the problem of ‘inheritance’ mentioned in §1.)
- Frege, G.** (1892) ‘Über Sinn und Bedeutung’, *Zeitschrift für Philosophie und philosophische Kritik* 100: 25-50; trans. ‘On Sense and Meaning’, in *Translations from the Philosophical Writings of Gottlob Frege*, trans. and ed. P.T. Geach and M. Black, Oxford: Blackwell, 3rd edn, 1980. (Introduces the notion of semantic presupposition. The page numbers of the original publication appear in the margins.)
- Stalnaker, R.C.** (1974) ‘Pragmatic Presupposition’, repr. in S. Davies, *Pragmatics: A Reader*, New York: Oxford University Press, 1991.
- Strawson, P.F.** (1950) ‘On Referring’, *Mind* 59: 320-44; repr. in *Logico-Linguistic Papers*, London: Methuen, 1971. (A classic paper, using presuppositional considerations to attempt to refute Russell’s theory of descriptions. Note that the term ‘presupposition’ is not used in the paper; it is called ‘implication in the special sense’.)
- Strawson, P.F.** (1952) *Introduction to Logical Theory*, London: Methuen. (Defines (Strawsonian) ‘presupposition’.)
- Strawson, P.F.** (1954) ‘A Reply to Mr. Sellars’, *Philosophical Review* 63: 216-31. (Extends presupposition beyond the assertoric.)

Probability, interpretations of

The term 'probability' and its cognates occur frequently in both everyday and philosophical discourse. Unlike many other concepts, it is unprofitable to view 'probability' as having a unique meaning. Instead, there exist a number of distinct, albeit related, concepts, of which we here mention five: the classical or equiprobable view, the relative frequency view, the subjectivist or personalist view, the propensity view, and the logical probability view. None of these captures all of our legitimate uses of the term 'probability', which range from the clearly subjective, as in our assessment of the likelihood of one football team beating another, through the inferential, as when one set of sentences lends a degree of inductive support to another sentence, to the obviously objective, as in the physical chance of a radioactive atom decaying in the next minute. It is often said that what all these interpretations have in common is that they are all described by the same simple mathematical theory - 'the theory of probability' to be found in most elementary probability textbooks - and it has traditionally been the task of any interpretation to conform to that theory. But this saying does not hold up under closer examination, and it is better to consider each approach as dealing with a separate subject matter, the structure of which determines the structure of the appropriate calculus.

1 The project

The task of interpreting probability might be approached in three distinct ways. The first way is to see the project as one of providing an explicit definition of the term 'probability' or, more usually, of the predicate 'has a probability of value p '. The second is to provide operational content to this predicate; in other words, to provide a set of procedures by means of which various probability values can be measured or attributed. This second task may result only in the provision of sufficient conditions for attributing the predicate, and conversely a solution to the first task may be accomplished without the definition providing us with a way to measure probability values. At one time, when verificationism held sway, a failure to provide measurement criteria was considered fatal to a definition (see [Logical positivism §4; Operationalism](#)), but we should keep separate these two tasks. The third approach is to provide an implicit definition of the predicate 'has a probability of value p ' by means of an axiomatized theory of probability, followed by an interpretation or model for that theory. That is, because the concept of probability is often considered to be mysterious and inaccessible to observation in a way that the concept of, say, blue, is not, the best and perhaps the only way of providing content to the concept is by constructing a detailed theory of probability and providing a model within which the theory is true. This third project consists in putting structural constraints on possible interpretations of a probability function, and it will not ordinarily result in an explicit definition. It will also not usually assign specific values to outcomes, except for certain extremal values, such as the certain or the impossible event.

2 The formal theory

The structure of elementary probability theory is often motivated by an appeal to familiar facts about relative frequencies. Suppose that a chance process, such as rolling a die, has N possible outcomes which form an outcome space $\Omega = \{1, 2, 3, 4, 5, 6\}$. By taking subsets of Ω , such as $\{2, 4, 6\}$, we have an event consisting in the die coming up even. Then take all possible subsets of Ω and this will form an algebra $\mathcal{A}$ of subsets of Ω , for example, a set of subsets of Ω closed under complementation ('negation') and union ('disjunction'). Let us say that an event E in $\mathcal{A}$ occurs if the elementary outcome, such as '2', is in E . Then we can define a probability function P over $\mathcal{A}$ by letting $P(E) = \frac{\text{the number of times } E \text{ occurs}}{\text{the total number of repetitions of the chance process}}$. This immediately gives us that

- (1) $P(\Omega) = 1$
- (2) $P(E) \geq 0$
- (3) If two events E and F are mutually exclusive, then $P(E \cup F) = P(E) + P(F)$.

For mathematical convenience, this elementary theory is usually extended by requiring countable additivity, wherein (3) is replaced by

- (3') $P(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} P(E_i)$ when the E_i are all mutually disjoint.

(The structure of $\mathcal{A}$ then needs to be more complex in ways that are inessential here.)

An important definition to add is that of conditional probability, $P(A | B)$, the probability of A given B , defined as $P(A | B) = P(A \cap B)/P(B)$. Two elements of $\mathcal{A}$ are *independent* if and only if $P(A | B) = P(A)$ or, equivalently, if $P(A \cap B) = P(A) \times P(B)$.

This abstract calculus is now no longer tied to the particular interpretation with which we began, and the algebra can be one of propositions rather than events, an ontology preferable for logical or subjective interpretations. Different accounts of probability can now be compared to this axiomatic theory.

3 Relative frequency theories

Here the ontology is one of event types, and the probability is explicitly defined either as the actual finite relative frequency as in §2 above, or as the limit value of the relative frequency when the total number of repetitions goes to infinity. This gives the value of the probability as an empirical property of the sequence or reference class of outcomes which generates the frequency, rather than absolutely. A naïve finite frequency interpretation is clearly unsatisfactory. Suppose a die is rolled a finite number of times N . Then no outcome, say a '6', can have a probability value more fine-grained than on a scale $0, 1/N, 2/N, \dots, 1$. Thus, after two throws, a '6' could only have a probability of 0, 0.5, or 1. This problem could be circumvented by requiring sufficiently many repetitions, where 'sufficiently many' would be a function of the number of possible outcomes and degree of precision required. A more standard remedy for this problem is to move to a limiting relative frequency interpretation, wherein the probability value is simply defined as the limit of m/N as $N \rightarrow \infty$, where m is the number of successes. This solves our first, definitional, task, but it produces a difficulty for the measurement project, for there is no guarantee that the value of the relative frequency after, say, 1,000 repetitions, will be the same as, or close to, the limiting value, because an unusual run of outcomes could occur initially. What one can do is appeal to the strong law of large numbers for binary valued outcomes, which asserts that if the repetitions of the experiment are independent and identically distributed (that is, the probability does not vary from repetition to repetition) then for every $\varepsilon > 0$, with probability one $|m/N - p| > \varepsilon$ only finitely often.

This provides assurance that in the long run, the probability that the limiting frequency will differ from the 'true' value of p is zero, but at the cost of introducing a second order probability which in turn needs to be interpreted. As an additional requirement, Richard von Mises (1957) correctly insisted that limiting frequency values can only be drawn from random sequences of data. Thus, if our die gave us the sequence of outcomes 1, 2, 3, 4, 5, 6, 1, 2, 3, 4, 5, 6, ..., the limiting frequency of a '2' would be 1/6, but quite clearly on every $(6N + 2)$ nd throw ($N = 0, 1, 2, \dots$) the probability of '2' would be 1, and would be 0 on any other throw. This highlights what is known as the problem of single-case probabilities - how do we transfer a probability value from a class to a single outcome? A standard answer is that the appropriate frequency for an outcome type is one drawn from a random sequence. Attempts to solve this single-case problem have led to complex and fascinating theories of randomness and statistical relevance, each of which is of philosophical interest in its own right (Sklar 1993).

4 Propensity theories

We have seen above that relative frequencies are defined relative to a class of outcomes. If we focus on the fact that outcomes can be produced by a fixed set of generating conditions, then it is reasonable to attribute probabilities to a system in a physical context, especially when the system producing the outcomes is irreducibly indeterministic. This gives us a chance disposition or *propensity* to produce a given outcome. For example, the probability of decay within the next minute is a physical property of a radioactive atom, just as is its atomic weight. This view, which seems to have originated with C.S. Peirce, and was resurrected by Karl Popper (see Popper, K.R. §3), has been criticized as excessively metaphysical. This is an unfair criticism, for the propensity value can be measured empirically for a system with a fixed propensity by employing the strong law of large numbers, mentioned above, together with statistical estimation techniques, thus satisfying the operational criterion. Being a thoroughly ontological interpretation, propensity accounts do not provide an explicit reductionist definition of probability, for propensities are basic, often primitive, properties of the world. However, no satisfactory solution has yet been given for the third project of providing a propensity calculus. Nor, despite a number of efforts by Popper and others, has a satisfactory argument been provided that quantum probabilities are

obviously propensities in any detailed sense (see [Causation](#); [Quantum mechanics, interpretation of](#); [Statistics](#)).

5 Subjective probabilities

In contrast to the objective interpretations just described, probability has always had a close relation with degrees of rational belief. Within this tradition, one can measure such degrees of belief operationally by means of betting behaviour. Here the elements to which the probabilities attribute values are propositions or sentences rather than events. By means of an ingenious operational process, known as the Dutch book method, within which the probability value is defined in terms of the lowest odds at which the gambler will accept a bet, one can show that plausible constraints on rational behaviour are satisfied if and only if axioms (1)–(3) and the definition of conditional probability of §2 hold. The subjective probability assignments are then said to be coherent. Thus we have satisfied the second and third criteria cited earlier, although it should be noted that different individuals can assign widely different values to a contingent proposition and both be coherent. De Finetti (1937), however, held that (3') was not true for subjective probabilities because placing an infinite number of bets makes no sense for human agents. The Dutch book method will give us what are called prior probabilities but, except for extremal values, these are not arrived at by an a priori process. Rather, they are the expression of an unarticulated amalgamation of background knowledge. Although there is some controversy over how best to revise beliefs in the light of empirical evidence, the traditional way to do this is via conditionalization, using Bayes' theorem. If H_j is a hypothesis (say that the die is fair), E is empirical evidence (say data from throws) and P_0 is our prior subjective probability assignment, then

$$P_1(H_j) = P_0(H_j \mid E) = P_0(E \mid H_j)P_0(H_j) / \sum_i P_0(E \mid H_i)P_0(H_i)$$

where $\{H_i\}$ is the set of hypotheses under consideration and P_1 is the new, posterior, probability assignment. Provided that the members of $\{H_i\}$ are distinct hypotheses (only one can be true at any time) and that $\sum_i P_0(H_i) = 1$ (the possibility that no member of $\{H_i\}$ is true is not entertained), $\sum_i P_0(E \mid H_i)P_0(H_i) = P_0(E)$, which gives $P_1(H_j) = P_0(H_j \mid E)P_0(H_j) / P_0(E)$.

The above gives us a normative theory of how agents should distribute their degrees of belief, but there has always been a divergence between this normative Bayesianism and individuals' actual degrees of belief. Investigations by psychologists and economists have revealed systematic differences from Bayesian prescriptions, even by agents well-versed in probabilistic reasoning.

6 Classical interpretations of probability

This approach allocates probability values by dividing the outcomes into equipossible cases, and then using a principle of indifference to give each equipossible case an equal probability. Thus, if one considers, on the basis of symmetry, that each side of a die is just as likely to come up as any other side, then the classical theory attributes a value of $1/6$ to each such outcome. This is not, contrary to many claims, always an a priori attribution, because it generally depends upon some specific empirical knowledge about symmetries of the system. The classical approach is unsatisfactory for a variety of reasons. First, it is limited to situations in which equipossible cases are available. Second, there are straightforward paradoxes associated with this approach, which allow different probability values to be attached to the same event. The simplest is one introduced by Bertrand. Suppose you have wine and water mixed, in a ratio somewhere between one part of water to one of wine, and two parts water to one of wine. Using an indifference principle on the ratio water:wine, we have that the probability of the ratio lying between 1 and 1.5 is $1/2$. Now consider the ratio wine:water. This can lie between $1/2$ and 1, and the water:wine ratio of $1/5$ is a wine:water ratio of $2/3$. The indifference principle then says that the probability of this ratio lying between 1 and $2/3$ has a probability of $2/3$. This is contradictory.

7 Logical interpretations and other approaches

The logical approach takes propositions as the objects to which probabilities are ascribed, and interprets a conditional probability as a logical relation giving a degree of inductive support from the conditioning sentence to the conditioned. Thus, $P(H \mid E)$ is the degree of inductive support that the evidence statement E gives to the hypothesis H . To measure $P(H \mid E)$, Carnap (1945) considered state descriptions. In the simplest case, one lists all the individuals $a_1 \dots a_n$ in a world and all the predicates $F_1 \dots F_r$. Then a state description is an attribution of

F_i or $\sim F_i$ to each individual a_j for all i and j , that is, a maximally consistent description of some possible world. If we now attribute, in an a priori way, a measure m on state descriptions (and this can be done in a number of ways), then the conditional logical probability is just $m(H \& E)/m(E)$. The principal drawback to using state descriptions is that learning from experience is impossible, simply because all predicates are logically independent and hence observing an instance of the property it represents gives no information about instances of any other. For this reason, Carnap switched to structure descriptions, within which individuals are indistinguishable. Interestingly, the differences between Maxwell-Boltzmann statistics, Bose-Einstein, and Fermi-Dirac statistics in physics are representable as differences that depend upon which states are physically possible. Because it is an empirical fact which particles satisfy which statistics, this sheds considerable doubt upon the whole enterprise of making a priori probability attributions.

There are now well-entrenched theories of comparative probability within which numerical values are not assigned, but one outcome is simply considered to be at least as probable as another. Alternatively, rather than assigning specific values to a proposition, intervals of probability can be assigned to mirror our uncertainty about the correct value via upper and lower probabilities. The connections between subjective probability and objective chance are now also the subject of much interest.

See also: Carnap, R. §5; Confirmation theory; Decision and game theory; Field theory, quantum; Inductive inference; Probability theory and epistemology; Rational beliefs; Rational choice theory; Reichenbach, H. §3

PAUL HUMPHREYS

References and further reading

- Carnap, R.** (1945) 'On Inductive Probability', *Philosophy of Science* 12: 72-97. (An accessible introduction to logical probability by its primary advocate.)
- Cohen, L.J.** (1989) *An Introduction to the Philosophy of Probability and Induction*, Oxford: Clarendon Press. (Chapter 2 is a brief but useful critical survey of the interpretations discussed in this article.)
- Feller, W.** (1968) *An Introduction to Probability Theory and Its Applications*, New York: John Wiley & Sons. (Still the best introduction to the mathematical theory of probability. Volume 1 treats discrete outcome spaces; volume 2 (2nd edition) 1971 deals with continuous outcome spaces.)
- Finetti, B. de** (1937) 'Foresight: its Logical Laws, its Subjective Sources'; repr. in H. Kyburg and H. Smokler (eds) *Studies in Subjective Probability*, New York: John Wiley & Sons, 1964. (A classic source on subjective probability.)
- Jeffrey, R.** (1983) *The Logic of Decision*, Chicago, IL: University of Chicago Press, 2nd edn. (A systematic exposition of one variety of Bayesianism.)
- Kahneman, D., Slovic, P. and Tversky, A.** (eds) (1982) *Judgement under Uncertainty: Heuristics and Biases*, Cambridge: Cambridge University Press. (A collection of papers by psychologists describing how individuals reason probabilistically. See §5.)
- Lewis, D.** (1980) 'A Subjectivist's Guide to Objective Chance', in R. Jeffrey (ed.) *Studies in Inductive Logic and Probability*, vol. 2, Berkeley, CA: University of California Press. (Reprinted with important Postscripts in D. Lewis, *Philosophical Papers*, vol. 2, Oxford: Oxford University Press, 1986. This article details some connections between objective and subjective probabilities.)
- Mises, R. von** (1957) *Probability, Statistics, and Truth*, New York: Dover, 2nd edn. (A nontechnical treatment of a relative frequency approach.)
- Popper, K.** (1974) 'Suppes's Criticisms of the Propensity Interpretation of Probability and Quantum Mechanics', in P.A. Schilpp (ed.) *The Philosophy of Karl Popper*, La Salle, IL: Open Court, 1, 125-39. (A detailed treatment of a propensity theory. The article to which it is replying - P. Suppes, 'Popper's Analysis of Probability in Quantum Mechanics', 760-74 - should be read too.)
- Sklar, L.** (1993) *Physics and Chance*, Cambridge: Cambridge University Press. (Chapter 3 is a readable overview of probability, with a section - §4 - on randomness.)

Proof theory

Proof theory is a branch of mathematical logic founded by David Hilbert around 1920 to pursue Hilbert's programme. The problems addressed by the programme had already been formulated, in some sense, at the turn of the century, for example, in Hilbert's famous address to the First International Congress of Mathematicians in Paris. They were closely connected to the set-theoretic foundations for analysis investigated by Cantor and Dedekind - in particular, to difficulties with the unrestricted notion of system or set; they were also related to the philosophical conflict with Kronecker on the very nature of mathematics. At that time, the central issue for Hilbert was the 'consistency of sets' in Cantor's sense. Hilbert suggested that the existence of consistent sets, for example, the set of real numbers, could be secured by proving the consistency of a suitable, characterizing axiom system, but indicated only vaguely how to give such proofs model-theoretically. Four years later, Hilbert departed radically from these indications and proposed a novel way of attacking the consistency problem for theories. This approach required, first of all, a strict formalization of mathematics together with logic; then, the syntactic configurations of the joint formalism would be considered as mathematical objects; finally, mathematical arguments would be used to show that contradictory formulas cannot be derived by the logical rules.

This two-pronged approach of developing substantial parts of mathematics in formal theories (set theory, second-order arithmetic, finite type theory and still others) and of proving their consistency (or the consistency of significant sub-theories) was sharpened in lectures beginning in 1917 and then pursued systematically in the 1920s by Hilbert and a group of collaborators including Paul Bernays, Wilhelm Ackermann and John von Neumann. In particular, the formalizability of analysis in a second-order theory was verified by Hilbert in those very early lectures. So it was possible to focus on the second prong, namely to establish the consistency of 'arithmetic' (second-order number theory and set theory) by elementary mathematical, 'finitist' means. This part of the task proved to be much more recalcitrant than expected, and only limited results were obtained. That the limitation was inevitable was explained in 1931 by Gödel's theorems; indeed, they refuted the attempt to establish consistency on a finitist basis - as soon as it was realized that finitist considerations could be carried out in a small fragment of first-order arithmetic. This led to the formulation of a general reductive programme.

Gentzen and Gödel made the first contributions to this programme by establishing the consistency of classical first-order arithmetic - Peano arithmetic (PA) - relative to intuitionistic arithmetic - Heyting arithmetic. In 1936 Gentzen proved the consistency of PA relative to a quantifier-free theory of arithmetic that included transfinite recursion up to the first epsilon number, ϵ_0 ; in his 1941 Yale lectures, Gödel proved the consistency of the same theory relative to a theory of computable functionals of finite type. These two fundamental theorems turned out to be most important for subsequent proof-theoretic work. Currently it is known how to analyse, in Gentzen's style, strong subsystems of second-order arithmetic and set theory. The first prong of proof-theoretic investigations, the actual formal development of parts of mathematics, has also been pursued - with a surprising result: the bulk of classical analysis can be developed in theories that are conservative over (fragments of) first-order arithmetic.

1 Metamathematics

Proof theory is a branch of mathematical logic founded by David Hilbert around 1920 to pursue his programme in the foundations of arithmetic (see [Hilbert's programme and formalism](#)). Consistency is the crucial logical notion connected with Hilbert's investigations. In traditional, Aristotelian logic it was viewed as a semantic notion: two or more statements are consistent if they are simultaneously true under some interpretation. In modern logic there is a syntactic definition that fits complex theories since Frege's *Begriffsschrift* (1879): a set of statements is consistent with respect to a logical calculus if no statement of the form $P \ \& \ \neg P$ is derivable from the statements by rules of the calculus. If these definitions are equivalent for a logic we have a significant fact, as the equivalence amounts to the completeness (and soundness) of the logic's system of rules. The first such completeness theorem was obtained for sentential logic by Paul Bernays (1918) and, independently, by Emil Post (1921); the completeness of predicate logic was proved by Kurt Gödel in 1930. The crucial step in such proofs shows that *syntactic* consistency implies *semantic* consistency. The converse is established quite directly, but involves the notion of truth. Here we have located one central issue that has motivated proof-theoretic investigations: namely, to avoid the uncritical use of the broad concept of classical truth for infinite mathematical structures.

Cantor applied consistency in an informal way to sets. He distinguished, for example in a letter to Dedekind, a consistent from an inconsistent multiplicity: the latter is such ‘that the assumption that all of its elements “are together” leads to a contradiction’, whereas the elements of the former ‘can be thought of without contradiction as “being together”’. Cantor had also conveyed these distinctions by letter to Hilbert in 1897. (See Purkert and Ilgauds (1987) for the text of the letters and Sieg (1990) for the wider historical context.) Hilbert pointed out, both implicitly (1900) and explicitly (1905), that Cantor had not given a rigorous criterion for distinguishing between consistent and inconsistent multiplicities. Hilbert suggested remedying the problem for analysis and giving the proof of the ‘existence of the totality of real numbers or - in the terminology of G. Cantor - the proof of the fact that the system of real numbers is a consistent (complete) set’ by establishing the consistency of an axiomatic characterization of the real numbers (1900). Indeed, he claimed that consistency could be established ‘by a suitable modification of familiar methods’. Hilbert’s own hints, partly in unpublished lecture notes, together with remarks by Bernays, make it plausible that he had a model-theoretic proof in mind. This problematic can be traced back to considerations in Dedekind’s work (1888) as explicated carefully in his letter to Keferstein, where he asked of his notion ‘simply infinite system’, ‘does such a system *exist* at all in the realms of our ideas? Without a logical proof of existence it would always remain doubtful whether the notion of such a system might not perhaps contain internal contradictions’. Clearly, Dedekind tried to establish consistency by exhibiting a suitable ‘logical’, that is, set-theoretic, model.

In 1904 Hilbert began to pursue a completely different strategy for giving consistency proofs. His new way of proceeding was still aimed at securing the existence of sets, but intended to exploit the formalizability of the theory at hand. Formalizations had to satisfy requirements stricter than those imposed on the structure of theories by the traditional (Euclidean) axiomatic-deductive method. The additional requirement was the regimentation of inferential steps in proofs: not only did axioms have to be given in advance, but the rules representing steps in mathematical arguments had to be taken from a predetermined list. To avoid a regression in the definition of proof and to achieve intersubjectivity at an absolutely minimal level, the rules had to be ‘formal’ or ‘mechanical’ and had to depend only on the syntactic form of statements. Thus, to exclude any ambiguity, a precise and effectively described language was also needed to formalize particular theories; the indications (Hilbert 1905) were very sketchy, however.

The general kind of requirements had been clear to Aristotle and were explicitly formulated by Leibniz; but only Frege, in his *Begriffsschrift*, presented - in addition to an expressively rich language with relations and quantifiers - an adequate logical calculus. Through the formalization of mathematical proofs Frege pursued a clear philosophical aim, namely to recognize the ‘epistemological nature’ of theorems. In the introduction to his *Grundgesetze der Arithmetik* (see Frege, G.), he wrote: ‘By insisting that the chains of inference do not have any gaps we succeed in bringing to light every axiom, assumption, hypothesis or whatever else you want to call it on which a proof rests; in this way we obtain a basis for judging the epistemological nature of the theorem.’ An epistemological analysis of theorems was also aimed for by Hilbert in his work on the foundations of geometry, but done quite differently, relying on a more traditional axiomatic presentation and pushing forward genuinely metamathematical investigations. For the emerging proof-theoretic work the formal aspect Frege had emphasized would be exploited by Hilbert in a distinctively novel way.

Hilbert repeatedly gave courses on the foundations of mathematics in the period 1904 to 1917; the spirit of these lectures is captured in his Zurich talk ‘*Axiomatisches Denken*’ (1918). The mathematical development and philosophical clarification of a new consistency programme began to be given only in Hilbert’s lectures on *Prinzipien der Mathematik* (Principles of Mathematics) presented in the winter term of 1917/1918. These lectures mark the beginning of the fruitful collaboration with Bernays, who supported Hilbert’s preparation in essential ways and also wrote careful notes. The notes of these and later lectures are amazing documents, as one finds in them for the first time a detailed modern presentation of the syntax and informal semantics of predicate logic and of finite type theories. The 1917/1918 notes served as the basis for Hilbert and Ackermann’s *Grundzüge der theoretischen Logik* (1928); indeed, with minor exceptions the notes contain all the material presented in the book. In lectures during the summer term of 1920 and the winter term of 1921/1922, proof theory and the ‘finitist consistency programme’ emerged.

2 Hilbert’s programme

For the purpose of the consistency programme, metamathematics was taken in Hilbert's 1921/1922 lectures to be included in, if not coextensive with, the part of mathematics acceptable to constructivists such as **Kronecker** and Brouwer. The point of consistency proofs was no longer to guarantee the existence of sets, but to establish the instrumental usefulness of classical mathematical theories T - say, set theory - with respect to finitist mathematics. That focus rested on the observation that the statement formulating the consistency of T is equivalent to the reflection principle

$$Pr(a, 'σ') \Rightarrow σ.$$

Here, Pr is the finitist proof predicate for T , $σ$ a finitistically meaningful statement, and ' $σ$ ' its translation into the language of T . A finitist consistency proof for T would thus ensure that T is a reliable instrument for the proof of finitist statements. Other important metamathematical issues were the completeness and decidability of theories.

The formalizability of mathematics was obviously crucial for this proof-theoretic approach, and the programmatic goal was seen as a way to circumvent some philosophical issues, for example, concerning the nature of infinite sets in the case of set theory. For these reasons Hilbert's philosophical position is (still) frequently equated with formalism in the sense of Frege's articles 'Über die Grundlagen der Geometrie' (1903, 1906; see **Frege, G.**) and of Brouwer's inaugural address 'Intuitionism and Formalism' (1913). Such a view is not completely unsupported by some of Hilbert's polemical remarks during the 1920s, but on balance his philosophical views developed into a sophisticated instrumentalism, if that label is taken in Ernest Nagel's judicious sense (see **Nagel 1961**). Hilbert's is an instrumentalism emphasizing the contentual motivation of mathematical theories; that is perhaps most clearly expressed in the first chapter of Hilbert and Bernays (1934). A sustained philosophical analysis of proof-theoretic research in the context of broader issues in the philosophy of mathematics was provided by Bernays; his penetrating essays stretch over five decades (see **Bernays 1976**). Feferman (1988) and Sieg (1988) give complementary accounts; for altogether contrasting approaches, see Detlefsen (1986) and Simpson (1988).

The mathematical work is most remarkable for what it started, as it constitutes the beginnings of modern mathematical logic. Even before the work of Gödel and Gentzen it was rich in accomplishments: consider, as examples, Bernays' completeness proof for sentential logic; the partial solutions to the decision problem (Hilbert's *Entscheidungsproblem*) for predicate logic obtained by Behmann, Bernays, Schönfinkel and Herbrand; and the consistency proofs given by Ackermann, von Neumann and Herbrand. Taking for granted the broader conceptual clarifications and the focus on first-order logic, Herbrand's '*théorème fondamentale*' is perhaps the most significant result on purely logical grounds, but also because of its applicability in consistency proofs. Gödel pointed to its essence when he gave in a 1933 lecture the following formulation of Herbrand's theorem: 'If we take a theory which is constructive in the sense that each existence assertion made in the axioms is covered by a construction, and if we add to this theory the non-constructive notion of existence and all the logical rules concerning it, for example, the law of excluded middle, we shall never get into any contradiction' (see **Herbrand's theorem**).

The results obtained in the 1920s, including Herbrand's theorem, were disappointing when measured against the hopes and ambitions of the Hilbert school: Ackermann, von Neumann and Herbrand had established essentially the consistency of first-order arithmetic with a very restricted principle of induction (for quantifier-free formulas). Actual limits on finitist considerations for consistency proofs had been reached; that became clear in 1931 through Gödel's theorems and the realization that finitist proofs could be formalized in a weak fragment of number theory (see **Gödel's theorems**). Initially, Gödel did not share the view on the limits of finitist reasoning, as is clear from the final remarks in his 'Über formal unentscheidbare Sätze der *Principia Mathematica* und verwandter Systeme I' (On Formally Undecidable Propositions of *Principia Mathematica* and Related Systems) (1931) and contemporaneous correspondence with von Neumann and Herbrand. But by December 1933, when he lectured in Cambridge, Massachusetts, he had changed his position: having isolated a 'system A' - essentially a version of primitive recursive arithmetic (PRA) - he made the following claim: 'Now all the intuitionistic [that is, finitist] proofs complying with the requirements of the system A which have ever been constructed can easily be expressed in the system of classical analysis and even in the system of classical arithmetic, and there are reasons for believing that this will hold for any proof which one will ever be able to construct.'

The solvability of the *Entscheidungsproblem* had been made implausible by Gödel's results, but the actual proof of unsolvability had to wait until 1936 for a conceptual clarification of 'mechanical procedure' or 'algorithm'. Such a

clarification was achieved mainly through the work of Church and Turing (see [Church's thesis](#); [Computability theory](#)). A precise notion of mechanical procedure was also needed to prove the incompleteness theorems for *general* 'formal' theories satisfying basic representability and derivability conditions; after all, Gödel had established limits only for (formalizations of) particular theories, such as the system of *Principia Mathematica* and then current axiomatic set theories. In his attempt to characterize a proper extension of the class of primitive recursive functions, Gödel introduced (in his Princeton lectures of 1934) the general recursive functions through an equational calculus (see [Computability theory §2](#)). The informal concept underlying Gödel's and also Church's approach - calculability of functions in a 'formal' calculus - was carefully analysed by Hilbert and Bernays ([1939: supplement 2](#)); they formulated recursiveness conditions for general deductive formalisms and showed that the number-theoretic functions whose values can be calculated in formalisms satisfying these conditions are exactly the general recursive ones.

The impact of the incompleteness theorems on Hilbert's programme was profound and yet limited. On the one hand, as remarked above, they pointed out definite limits of finitist considerations; on the other hand, they left open the possibility of modifying the programme to a *general reductive programme* that was no longer aiming for 'absolute' finitist consistency proofs, but rather for consistency proofs relative to 'appropriate' constructive theories. Before Gödel's results were known, Bernays ([1930](#)) had given a detailed and searching analysis of the philosophical aims of Hilbert's proof theory; in the *Postscriptum* published in 1976, Bernays expressed clearly what was lost due to the incompleteness theorems: the sharp distinction between what is intuitive and what is non-intuitive, a distinction that was basic for the proposed philosophical treatment of the problem of the infinite. Thus, it is this particular 'solution' to a philosophical problem that was shown to be impossible.

Work in proof theory continued with the explicit goal of achieving relative consistency proofs. Such work is in a venerable mathematical tradition, as the many examples of significant results show: for example, the consistency of non-Euclidean relative to Euclidean geometry; that of Euclidean geometry relative to analysis; the consistency of set theory with the axiom of choice relative to set theory without it; that of set theory with the negation of the axiom of choice relative to set theory. The mathematical significance of relative consistency proofs is often brought out by sharpening them to conservative extension results. Such results may ensure, for example, that the theories have the same class of provably total functions (see §6). However, the initial motivation for such arguments is most frequently philosophical: one wants to guarantee the coherence of the original theory on an epistemologically distinguished basis. One has to see the specific results that have been obtained in the pursuit of the general reductive programme from this perspective.

3 Mathematical work

The development of the general reductive programme is characterized by modifications of the two-pronged approach of Hilbert's original programme, namely (1) weakening the theories in which parts of mathematics are formalized, and (2) strengthening the theories in which the metamathematical considerations are carried out and, consequently, relative to which constructive consistency proofs can be given. The first modification reaches back to Weyl's *Das Kontinuum* (1918) and culminated in the 1970s, when it was realized that the classical results of mathematical analysis can be obtained in conservative extensions of first-order arithmetic. The second modification started with the work of Gödel and Gentzen in 1933, when they established independently the consistency of classical arithmetic relative to intuitionistic arithmetic; it led in the 1970s and 1980s to consistency proofs of subsystems of second-order arithmetic or, synonymously, subsystems of analysis relative to intuitionistic theories of constructive ordinals. Obviously, only a sketch of some main results can be attempted here.

Second-order arithmetic has been used for some time as a framework for the formal development of classical mathematical analysis; that is, the theory of the continuum set-theoretically described by Dedekind and Cantor. Because of this mathematical adequacy, second-order arithmetic and a variety of subsystems have been thoroughly investigated. The main set-theoretic principles for these systems are the comprehension axiom (CA)

$$(\exists X)(\forall y)(y \in X \Leftrightarrow S(y))$$

and the axiom of choice (AC) in the form

$$(\forall x)(\exists Y)S(x, Y) \Rightarrow (\exists Z)(\forall x)S(x, Z_x),$$

where S is in each case an arbitrary formula of the language and thus may contain set quantifiers; $y \in Z_x$ is defined to mean $\langle y, x \rangle \in Z$. These principles are impredicative, as the sets X and Z whose existence is postulated are in general characterized by reference to all sets of natural numbers. The induction principle is formulated either as a schema or as a second-order axiom

$$(\forall X)(0 \in X \ \& \ (\forall y)(y \in X \Rightarrow y' \in X) \\ \Rightarrow (\forall x)x \in X).$$

Theories are denoted by the name of their set-existence principle enclosed in parentheses; thus (CA) names full analysis. If $\downarrow$ follows its name, a theory uses the second-order axiom to formalize induction. Two general results are of interest, as they show that second-order arithmetic has a certain robustness: (CA) is proof-theoretically equivalent to Zermelo-Fraenkel set theory without the power-set axiom; (AC) is conservative over (CA) for Π_4^1 -formulas and properly stronger, as there is a Π_2^1 -instance of AC not provable in (CA). In the presence of full CA, the theories with the induction schema, respectively, the second-order axiom of induction are equivalent, but they can be of strikingly different strength, when the set-existence principles are restricted; for example, $(\Pi_\infty^0\text{-CA})\downarrow$ is conservative over Peano arithmetic (PA), whereas $(\Pi_\infty^0\text{-CA})$ proves the consistency of PA.

In 1917/1918 Hilbert and Bernays used ramified type theory with the axiom of reducibility to develop analysis; the published presentation (1939: supplement 4) is based on this early work, but employs full second-order arithmetic as the formal framework. They encouraged developments with restricted means already in 1920, writing of Brouwer and Weyl:

The positive and fruitful part of the investigations into the foundations of mathematics carried out by these two researchers fits into the mould of the axiomatic method and is exactly in the spirit of this method. For one investigates here, how a part of analysis can be delimited by a certain narrower system of assumptions.

(Hilbert and Bernays 1920: 34)

Subsystems of analysis are now mainly defined by restricting S in the set-existence schemata to particular classes of formulas. The set-theoretic demands can be reduced dramatically: Hilbert and Bernays' presentation (1939), for example, can be given quite readily in $(\Pi_1^1\text{-CA})\downarrow$. Strictly mathematical work continued to accompany work on consistency proofs for subsystems; it had the aim of establishing the mathematical significance of subsystems and made use of work in the constructivist tradition.

By the mid-1970s, through final efforts of Feferman, Friedman and Takeuti, it was clear that classical analysis could be carried out in conservative extensions of number theory, for example, $(\Pi_\infty^0\text{-CA})\downarrow$. In this context Friedman suggested pursuing a strategy familiar from investigations of the axiom of choice in set theory; namely, to establish the equivalence of certain set-existence principles with mathematical theorems. This theme is played with surprising variations in Friedman's and Simpson's work on subsystems and gave rise to the enterprise of 'reverse mathematics'. To mention just two, by now classic, examples: CA for arithmetic formulas is equivalent to the theorem that every bounded sequence of reals has a least upper bound and to König's lemma. Friedman introduced a second-order theory WKL_0 that extends primitive recursive arithmetic conservatively for Π_2^0 -formulas; the theory is weak, but still provides a very good basis for developing parts of analysis and algebra. Simpson (1988) not only considers the development of mathematics in such a conservative extension of PRA a 'reductionist programme', but equates it with Hilbert's programme. One should recall, however, that Hilbert did not propose to redo all of mathematics in (a conservative extension of) PRA, but rather to justify - via finitist consistency proofs - the use of strong classical theories sufficient for the direct formalization of mathematical practice.

4 Logical tools

Hilbert's central idea for the metamathematical treatment of the consistency problem found its expression in what is known as the ε -calculus and the associated substitution method. Hilbert and Bernays (1939) present what was achieved in its terms; even Herbrand's (difficult) work was recast in terms of the ε -calculus. This tradition was kept up by Tait (1965) and more recently by Mints (1994). However, other logical tools turned out to be more useful for proof-theoretic investigations: Gentzen's sequent calculi (see [Natural deduction, tableau and sequent systems §2](#)) and Gödel's so-called *Dialectica* interpretation. In the latter, Gödel used computable functionals of

finite type to obtain a reduction of intuitionistic arithmetic (HA); joining this reduction with the consistency of PA relative to HA, a consistency proof relative to the system of these functionals was obtained. Influenced by the considerations of Hilbert (1926), Gödel presented this work in a lecture at Yale in 1941, but published it only in 1958 (see [Gödel's theorems](#)). Spector (1962) gave a consistency proof for full classical analysis using bar recursive functionals of finite type; this proof prompted a searching analysis in the *Stanford Report on the Foundations of Analysis* (1963).

Here the focus is on sequent calculi, as they have found the most extensive use and widest applicability. In the form given to them by Tait (1968), they allow the proof of finite sets of formulas built up from literals (atomic formulas or negations of such), conjunction, disjunction, universal and existential quantification and - depending on the theory - infinitary conjunction and disjunction. Thus, for just first-order logic the basic logical symbols are ‘ $\wedge$ ’, ‘ $\vee$ ’, ‘ $\exists$ ’, ‘ $\forall$ ’ and the rules of the calculi include the following ones, where Γ is used as a syntactic variable ranging over finite sets of formulas, Γ, ϕ stands for the union of Γ and the singleton ϕ , and $a \in P(\Gamma)$ means that the parameter a occurs in one of the formulas in Γ :

LA: $\Gamma, \varphi, \neg\varphi, \quad \varphi \text{ atomic}$

$\wedge$:
$$\frac{\Gamma, \varphi_0 \quad \Gamma, \varphi_1}{\Gamma, \varphi_0 \wedge \varphi_1}$$

$\vee_i$:
$$\frac{\Gamma, \varphi_i}{\Gamma, \varphi_0 \vee \varphi_1}, \quad i = 0, 1$$

C :
$$\frac{\Gamma, \varphi \quad \Gamma, \neg\varphi}{\Gamma}$$

$\forall$:
$$\frac{\Gamma, \varphi a}{\Gamma, (\forall x)\varphi x} \quad a \notin P(\Gamma)$$

$\exists$:
$$\frac{\Gamma, \varphi t}{\Gamma, (\exists x)\varphi x}$$

The crucial claim established by Gentzen was his ‘*Hauptsatz*’ or cut-elimination theorem: every derivation in the logical system using the cut rule C can be transformed into a cut-free or normal derivation. Inspecting the rules, one notices that the premises of all the rules (except for C) contain only subformulas of formulas in the conclusion. Consequently, a normal derivation of a sequent Γ contains only subformulas of elements in Γ . This is the crucial subformula property of normal derivations, providing a bound on the complexity of formulas that can occur in a proof of Γ . In Gentzen (1935) this metamathematical fact is established and used to obtain in a most perspicuous way the most far-reaching consistency result that had then been obtained, namely Herbrand’s.

For full first-order number theory PA this treatment was extended by Gentzen (1936) to a partial cut-elimination argument whose termination was established by quantifier-free transfinite induction up to the first epsilon number - $\text{TI}(\varepsilon_0)$. Gentzen showed (1943) that this induction schema for every ordinal α less than ε_0 can be established in PA; this is the first ‘ordinal analysis’ of a formal theory. In the 1950s Lorenzen and, much more extensively, Schütte used infinitary extensions of Gentzen’s finitary systems; in particular for the treatment of PA they used the so-called ω -rule that allows one to infer $\Gamma, (\forall x)\phi x$ from the premises $\Gamma, \phi n$ for each natural number n (see [Non-constructive rules of inference §1](#)). Though derivations are now infinite, PA-derivations can be embedded into finitistically described ones and can be transformed effectively into cut-free derivations; the natural ordinal length of these derivations is bounded by ε_0 . Schütte extended these methods to treat systems of ramified analysis RA_α (of order α) and obtained, in particular, ordinal bounds on the length of normal derivations in terms of the Veblen hierarchy of ordinal functions. This work was used in 1963 by Feferman and Schütte independently to characterize the ordinal Γ_0 of predicative analysis, that is, the first ordinal α such that $\text{TI}(\alpha)$ cannot be proved in RA_β for β less than α .

5 Reductive results

The mathematical work described in §3 was to a large extent inspired by the attempt to establish the significance of relative consistency proofs and/or to focus on manageable subsystems that might be a proper next target for such proofs. Most of the reductive results mentioned below can be established by the metamathematical tool of the sequent calculi, though some of the original proofs used different techniques.

The Gödel-Gentzen result for number theory can be extended to ramified analysis and also to obtain conservativeness for Π_2^0 -formulas. This provides a relative consistency proof, as ramified systems with intuitionistic logic are certainly acceptable constructively (as long as the ordinals along which the systems are iterated are acceptable). In the early 1960s, partly through the study of predicativity, significant subsystems with S restricted to small classes of analytic formulas were isolated, among them $(\Sigma_1^1\text{-AC})$ and $(\Delta_1^1\text{-CA})$. Kreisel pointed out that

$$(\Sigma_1^1\text{-AC}) \supseteq (\Delta_1^1\text{-CA}).$$

Friedman showed that $(\Sigma_1^1\text{-AC})$ is conservative over $(\Delta_1^1\text{-CA})$ for Π_2^1 -formulas; that the inclusion is proper was established later by Steel. Indeed, $(\Sigma_1^1\text{-AC})$ is conservative for Π_2^1 -formulas over $(\Pi_0^1\text{-CA})_{<\varepsilon_0}$, a theory based on the transfinite iteration of the jump-operator and equivalent to ramified analysis of level less than ε_0 . This result, due to Friedman, allowed the determination of the proof-theoretic ordinal of the systems, but it also showed - and this was quite unexpected - that these *prima facie* impredicative theories were predicative in the sense of Feferman and Schütte. Feferman (1964) and Kreisel (1968) provide excellent summaries.

Friedman's theorem for $(\Sigma_1^1\text{-AC})$ turned out to be a special case of a general result: $(\Sigma_{n+1}^1\text{-AC})$ is conservative over $(\Pi_n^1\text{-CA})_{<\varepsilon_0}$ for formulas in F_n , where F_0 is Π_2^1 , F_1 is Π_3^1 , and F_n is Π_4^1 for $n > 1$ (see Friedman 1970). As just explained, the case $n = 0$ was of special interest for the study of predicativity; the case $n = 1$ is also to be placed in the context of foundational investigations. The most immediate context is provided by Feferman (1970), who relates the systems $(\Pi_1^1\text{-CA})_{<\varepsilon_0}$ to the classical theory for (less than ε_0 times iterated) classes defined by generalized inductive definitions. Well-known examples are the classes O of constructive ordinals and W of recursive well-founded trees. Between them, Friedman and Feferman reduced the subsystems $(\Sigma_2^1\text{-AC})$ and $(\Delta_2^1\text{-CA})$ to the classical theory of the tree classes $\mathbf{W}_\nu$ with index ν less than ε_0 . Feferman (1977) gives a detailed survey of mathematical and proof-theoretic work, including these last results.

Kreisel had introduced intuitionistic theories of iterated inductive definitions in the *Stanford Report* (1963); these theories were viewed as codifying constructive principles that might be used in consistency proofs for subsystems of analysis. Feferman's and Friedman's work described above established connections between subsystems of analysis and classical theories of inductive definitions, making a crucial step towards answering the major problem posed by Kreisel (1968): reduce $(\Sigma_2^1\text{-AC})$ to a constructive theory of inductive definitions. That would provide, as Kreisel put it, 'a solution to Hilbert's problem for the subsystem of analysis... $(\Sigma_2^1\text{-AC})$ '. The classical theories for classes defined by generalized inductive definitions are reducible to intuitionistic theories for accessible such classes. This allowed, in particular, a satisfactory solution of Kreisel's open problem, as $(\Sigma_2^1\text{-AC})$ was shown to be reducible to $(\text{ID})_{<\varepsilon_0}(\mathbf{O})$, the intuitionistic theory of constructive number classes with index less than ε_0 . These and many further results were obtained by Buchholz *et al.* (1981); in particular, the ordinal analysis of the subsystems at hand. This work was influenced by earlier considerations of Howard, Tait and Takeuti.

Subsystems of set theory, in particular of admissible set theory, were used by Jäger and Pohlers in the early 1980s to provide a unifying approach to the investigations and, it was hoped, an avenue for analysing even stronger systems than those corresponding to the subsystems mentioned above (see Jäger 1986). This has indeed been a successful strategy and reveals, in the work of Rathjen and others, a deep connection between large cardinals and the constructive ordinals needed for the proof-theoretic investigation of such systems. Rathjen (1995) succeeded in analysing $(\Pi_2^1\text{-CA})$; it seems that the techniques developed for this case might allow the treatment of full analysis. These proof-theoretic investigations can no longer be motivated by the concern of 'securing' mathematical practice: the systems that are investigated are much stronger than needed for practice; the constructive ordinals used in the metamathematical theory are obtained in analogy to large cardinals in set theory.

The systems WKL_0 , $(\Pi_\infty^0\text{-CA})|$ and $(\Pi_1^1\text{-CA})|$ have been recognized as significant for the formalization of

mathematical practice and are reducible to theories based on principles that are acceptable from constructive positions; after all, they are conservative for Π_2^0 -formulas over PRA, HA and the intuitionistic theory for the finite constructive number classes $(ID)_{<\omega}(\mathbf{O})$. This provides a coherent perspective bringing out the complementary character of mathematical and metamathematical work that ultimately aims at relating significant parts of mathematical practice to distinctive foundational positions. But there is no obvious answer to the question ‘What is the mathematical significance of those subsystems, when taken as vehicles for the formal axiomatic study of ordinary mathematics?’; similarly there is no obvious answer to the question ‘What is the philosophical significance of the corresponding systems PRA, HA and $(ID)_{<\omega}(\mathbf{O})$, when taken as formal expressions of foundational positions?’. There is ample room for reflection on these questions; the work reported here and in the literature provides rich and crucial data (see [Sieg 1990](#)). For the philosophical reflection on the foundations of mathematics the investigations of subsystems of set theory provide additional, significant material: what are constructions that lead to ‘accessible domains’ and how is it that we recognize their associated laws?

6 Outlook

‘Internal’ mathematical and philosophical challenges of work in proof theory were sketched at the end of the previous section. However, the foundational goals of proof-theoretic investigations have been complemented over the last few decades by other important initiatives.

First, in the 1950s, Kreisel initiated work that was to exploit the gap between provability in particular formal theories and truth. That led, on the one hand, to ‘global’ characterizations of the provably total functions of theories and to related independence results. On the other hand, by attending ‘locally’ to mathematical details of proofs and by using proof-theoretic techniques, it led to explicit computational information of mathematical significance. This seems to have come to fruition through work by Luckhardt ([1989](#)) and Kohlenbach ([1995](#)).

Second, methods and results of mathematical logic, but in particular of proof theory, have played an increasing role in computer science. Clearly, there has also been significant and stimulating influence in the other direction. Various (type) systems - for example, Martin-Löf’s, Girard’s system F, and Feferman’s systems of explicit mathematics - have been used for the presentation of proofs and computations, but also for describing transformations on them.

Third, there is a direct connection to the general topic of theorem proving; investigations here, when focusing on automated proof search, might reflect back into proof theory by providing data for a ‘structural theory of (mathematical) proofs’. Such a structural proof theory would go beyond the representation of proofs in formal theories and articulate search heuristics expressing leading mathematical ideas for particular parts of mathematics. Saunders Mac Lane suggested such an extension of proof-theoretic investigations in his Göttingen dissertation of 1934, but it has been pursued only recently through computer implementation of appropriate search procedures.

See also: [Hilbert’s programme and formalism](#); [Logical and mathematical terms, glossary of](#)

WILFRIED SIEG

References and further reading

- Bernays, P.** (1918) ‘Beiträge zur axiomatischen Behandlung des Logik-Kalküls’ (Contributions to the Axiomatic Treatment of the Logical Calculus), unpublished *Habilitationsschrift*, Göttingen University; published in much revised form, ‘Axiomatische Untersuchung des Aussagen-Kalküls der *Principia Mathematica*’ (Axiomatic Investigation of the Propositional Calculus of *Principia Mathematica*), *Mathematische Zeitschrift* 25: 305-20, 1926. (The first proof of the completeness of a calculus for sentential logic.)
- Bernays, P.** (1930) ‘Die Philosophie der Mathematik und die Hilbertsche Beweistheorie’ (Philosophy of Mathematics and Hilbert’s Proof Theory), *Blätter für Deutsche Philosophie* 4: 326-67; repr. in *Abhandlungen zur Philosophie der Mathematik (Essays on the Philosophy of Mathematics)*, Darmstadt: Wissenschaftliche Buchgesellschaft, 1976. (A searching analysis of the background for and the goals of Hilbert’s proof theory.)
- Bernays, P.** (1976) *Abhandlungen zur Philosophie der Mathematik (Essays on the Philosophy of Mathematics)*, Darmstadt: Wissenschaftliche Buchgesellschaft. (Five decades of Bernays’ essays.)
- Brouwer, L.E.J.** (1913) ‘Intuitionism and Formalism’, *Bulletin of the American Mathematical Society* 20: 81-96; repr. in P. Benacerraf and H. Putnam (eds) *Philosophy of Mathematics: Selected Readings*, Cambridge: Cambridge University Press, 2nd edn, 1983, 77-89. (A formalism with which Hilbert’s philosophical position is

frequently equated.)

- Buchholz, W., Feferman, S., Pohlers, W. and Sieg, W.** (1981) *Iterated Inductive Definitions and Subsystems of Analysis*, Berlin, Heidelberg and New York: Springer. (First proof-theoretic treatment of some impredicative subsystems of analysis.)
- Davis, M.** (ed.) (1965) *The Undecidable: Basic Papers on Undecidable Propositions, Unsolvability Problems and Computable Functions*, Hewlett, NY: Raven Press. (Anthology of the fundamental papers on the subject by Gödel, Church, Turing, Kleene, Rosser and Post.)
- Dedekind, R.** (1888) *Was sind und was sollen die Zahlen?*, Braunschweig: Vieweg; trans. W.W. Beman (1901), 'The Nature and Meaning of Numbers', in *Essays on the Theory of Numbers*, New York: Dover, 1963. (The modern set-theoretic foundation of number theory.)
- Detlefsen, M.** (1986) *Hilbert's Program: An Essay on Mathematical Instrumentalism*, Boston, MA, and Dordrecht: Reidel. (A contrasting approach to those of Feferman (1988) and Sieg (1988).)
- Feferman, S.** (1964) 'Systems of Predicative Analysis', *Journal of Symbolic Logic* 29: 1-30. (An excellent summary of work described in §5.)
- Feferman, S.** (1970) 'Formal Theories for Transfinite Iterations of Generalized Inductive Definitions and Some Subsystems of Analysis', in J. Myhill, R.E. Vesley and A. Kino (eds) *Intuitionism and Proof Theory*, Amsterdam: North Holland. (See §5.)
- Feferman, S.** (1977) 'Theories of Finite Type Related to Mathematical Practice', in J. Barwise (ed.) *Handbook of Mathematical Logic*, Amsterdam: North Holland. (A detailed survey of proof-theoretic work.)
- Feferman, S.** (1988) 'Hilbert's Program Relativized: Proof-Theoretical and Foundational Reductions', *Journal of Symbolic Logic* 53: 364-84. (A philosophical analysis of proof-theoretic research.)
- Frege, G.** (1879) *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*, Halle: Nebert; trans. 'Begriffsschrift, a Formula Language, Modelled Upon That of Arithmetic, for Pure Thought', in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967, 1-82. (This small booklet contains the most significant step from Aristotelian to modern logic.)
- Friedman, H.** (1970) 'Iterated Inductive Definitions and Σ_2^1 -AC', in J. Myhill, R.E. Vesley and A. Kino (eds) *Intuitionism and Proof Theory*, Amsterdam: North Holland. (See §5.)
- Gentzen, G.** (1935) 'Untersuchungen über das logische Schließen', *Mathematische Zeitschrift* 39: 176-210, 405-565; repr. Darmstadt: Wissenschaftliche Buchgesellschaft, 1969; trans. 'Investigations into Logical Deduction', *American Philosophical Quarterly* 1: 288-306, 2: 204-18, 1964; repr. in *The Collected Papers of Gerhard Gentzen*, ed. M.E. Szabo, Amsterdam and London: North Holland, 1969, 68-131. (Gentzen's doctoral dissertation, including the calculi of natural deduction and sequents, and the consistency proof for arithmetic with quantifier-free complete induction.)
- Gentzen, G.** (1936) 'Die Widerspruchsfreiheit der reinen Zahlentheorie', *Mathematische Annalen* 112: 493- 565; repr. Darmstadt: Wissenschaftliche Buchgesellschaft, 1967; trans. 'The Consistency of Elementary Number Theory', in *The Collected Papers of Gerhard Gentzen*, ed. M.E. Szabo, Amsterdam and London: North Holland, 1969, 132-213. (See §4.)
- Gentzen, G.** (1943) 'Beweisbarkeit und Unbeweisbarkeit von Anfangsfällen der transfiniten Induktion in der reinen Zahlentheorie', *Mathematische Annalen* 119 (1): 140-61; trans. 'Provability and Nonprovability of Restricted Transfinite Induction in Elementary Number Theory', in *The Collected Papers of Gerhard Gentzen*, ed. M.E. Szabo, Amsterdam and London: North Holland, 1969, 287-308. (The first 'ordinal analysis' of a formal theory; see §4.)
- Gödel, K.** (1986-95) *Kurt Gödel: Collected Works*, ed. S. Feferman et al., New York and Oxford: Oxford University Press, 3 vols. (Includes all the individual papers mentioned in the text.)
- Hilbert, D.** (1900) 'Über den Zahlbegriff' (On the Concept of Number), *Jahresbericht der Deutschen Mathematiker-Vereinigung* 8: 180-94; revised as appendix 6 of *Die Grundlagen der Geometrie*, Leipzig and Berlin: Teubner, 7th edn, 1930; trans. *The Foundations of Geometry*, Chicago, IL: Open Court, 2nd edn, 1971. (Hilbert's first paper addressing the foundations of arithmetic.)
- Hilbert, D.** (1905) 'Über die Grundlagen der Logik und der Arithmetik', in A. Krazier (ed.) *Verhandlungen des dritten internationalen Mathematiker-Kongresses in Heidelberg vom 8. bis 13. August 1904*, Leipzig: Teubner, 174-85; trans. 'On the Foundations of Logic and Arithmetic', in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967, 130-8. (Discusses the need for a greater formalization of proof theory.)
- Hilbert, D.** (1918) 'Axiomatisches Denken' (Axiomatic Thinking), *Mathematische Annalen* 78: 405-15; repr. in *Gesammelte Abhandlungen*, vol. 3, Berlin: Springer, 1935. (Captures the spirit of Hilbert's lecture courses on

the foundations of mathematics.)

- Hilbert, D.** (1920) 'Probleme der mathematischen Logik' (Problems of Mathematical Logic), lecture notes available in the Hilbert Nachlaß at the University of Göttingen, 46 pages. (Lecture notes from the summer term of 1920, written by N. Schönfinkel and P. Bernays.)
- Hilbert, D.** (1926) 'Über das Unendliche', *Mathematische Annalen* 95: 161-90; trans. 'On the Infinite', in P. Benacerraf and H. Putnam (eds) *Philosophy of Mathematics: Selected Readings*, Cambridge: Cambridge University Press, 2nd edn, 1983. (Fully developed formulation of the consistency programme and finitist standpoint; 'proof' of Cantor's continuum theorem.)
- Hilbert, D. and Ackermann, W.** (1928) *Grundzüge der theoretischen Logik (Principles of Theoretical Logic)*, Berlin: Springer, 2nd edn, 1938; trans. L.M. Hammond, G.G. Leckie and F. Steinhardt, *Principles of Mathematical Logic*, ed. R.E. Luce, New York: Chelsea, 1950. (Based on Hilbert's lectures in the winter term of 1917/1918.)
- Hilbert, D. and Bernays, P.** (1934, 1939) *Grundlagen der Mathematik (Foundations of Mathematics)*, Berlin: Springer, 2 vols; 2nd edn, 1968, 1970. (Classic work of mathematical logic; includes valuable statements of Hilbert's conception of the axiomatic method and finitary reasoning.)
- Jäger, G.** (1986) *Theories for Admissible Sets: A Unifying Approach to Proof Theory*, Naples: Bibliopolis. (Uses subsystems of set theory to provide a unifying approach to the investigation of consistency proofs.)
- Kohlenbach, U.** (1995) *Real Growth in Standard Parts of Analysis, Habilitationsschrift*, Frankfurt University. (Develops work initiated by Kreisel, exploiting the gap between provability and truth; see §6.)
- Kreisel, G.** (ed.) (1963) *Stanford Report on the Foundations of Analysis*, unpublished seminar report, Stanford University. (Important papers (some published elsewhere) on subsystems of analysis that emerged from a seminar held at Stanford in the summer of 1963; the contributors are J. Harrison, W.A. Howard, G. Kreisel, R. Parikh and W.W. Tait.)
- Kreisel, G.** (1968) 'A Survey of Proof Theory', *Journal of Symbolic Logic* 33: 321-88. (An excellent summary of work described in §5.)
- Luckhardt, H.** (1989) 'Herbrand Analysen zweier Beweise des Satzes von Roth: Polynomiale Anzahlschranken' (Herbrand Analyses of Two Proofs of Roth's Theorem: Polynomial Bounds), *Journal of Symbolic Logic* 54: 234-63. (Develops work initiated by Kreisel, exploiting the gap between provability and truth; see §6.)
- Mints, G.** (1994) 'Gentzen-Type Systems and Hilbert's Epsilon Substitution Method I', in D. Prawitz, B. Skyrms and D. Westerståhl (eds) *Proceedings of the Ninth International Conference, Uppsala, Sweden, 1991*, vol. 9, *Logic, Methodology and Philosophy of Science*, Amsterdam: North Holland. (Continues use of the ε -calculus used by Hilbert; see §4.)
- Nagel, E.** (1961) *The Structure of Science - Problems in the Logic of Scientific Explanation*, New York: Harcourt Brace. (Hilbert's views developed into instrumentalism in Nagel's sense.)
- Post, E.** (1921) 'Introduction to a General Theory of Elementary Propositions', *American Journal of Mathematics* 43: 163-85; repr. in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematics, 1879-1931*, Cambridge, MA: Harvard University Press, 1967, 264-83. (Proves the completeness of a calculus for sentential logic.)
- Purkert, W. and Ilgauds, H.J.** (1987) *Georg Cantor 1845-1918*, Basle: Birkhäuser. (A biography of Cantor that includes a number of very informative letters from Cantor to Hilbert from the late 1890s.)
- Rathjen, M.** (1995) 'Recent Advances in Ordinal Analysis: $(\Pi_2^1\text{-CA})$ and Related Systems', *Bulletin of Symbolic Logic* 1: 468-85. (Analysis of $(\Pi_2^1\text{-CA})$.)
- Schwichtenberg, H.** (1977) 'Proof Theory: Some Applications of Cut-Elimination', in J. Barwise (ed.) *Handbook of Mathematical Logic*, Amsterdam: North Holland, 867-95. (An elegant presentation of the proof-theoretic analysis of elementary number theory, PA.)
- Sieg, W.** (1988) 'Hilbert's Program Sixty Years Later', *Journal of Symbolic Logic* 53: 338-48. (A philosophical analysis of proof-theoretic research.)
- Sieg, W.** (1990) 'Relative Consistency and Accessible Domains', *Synthese* 84: 259-97. (Provides a historical and philosophical perspective on current proof-theoretic work.)
- Simpson, S.** (1988) 'Partial Realization of Hilbert's Program', *Journal of Symbolic Logic* 53: 349-63. (A contrasting approach to those of Feferman (1988) and Sieg (1988).)
- Spector, C.** (1962) 'Provably Recursive Functionals of Analysis: A Consistency Proof of Analysis by an Extension of Principles Formulated in Current Intuitionistic Mathematics', in *Recursive Function Theory*, Proceedings of Symposia of Pure Mathematics, Providence, RI: American Mathematical Society. (A consistency proof for full classical analysis.)

- Tait, W.W.** (1965) 'The Substitution Method', *Journal of Symbolic Logic* 30: 175-92.(Continues use of the ε -calculus used by Hilbert; see §4.)
- Tait, W.W.** (1968) 'Normal Derivability in Classical Logic', in J. Barwise (ed.) *The Syntax and Semantics of Infinitary Languages*, New York: Springer, 204-36.(Introduces and uses an elegant and technically convenient formulation of sequent calculi for the investigation of classical logic and subsystems of analysis.)
- Weyl, H.** (1918) *Das Kontinuum: Kritische Untersuchungen über die Grundlagen der Analysis*, Leipzig: Veit; trans. T. Bole, *The Continuum: A Critical Examination of the Foundations of Analysis*, New York: Dover, 1994.(The first modification of Hilbert's original programme.)

Proper names

The Roman general Julius Caesar was assassinated on 14 March 44 BC by conspirators led by Brutus and Cassius. It is a remarkable fact that, in so informing or reminding the reader, the proper names 'Julius Caesar', 'Brutus' and 'Cassius' are used to refer to three people each of whom has been dead for about two thousand years. Our eyes could not be used to see any of them, nor our voices to talk to them, yet we can refer to them with our words.

The central philosophical issue about proper names is how this sort of thing is possible: what exactly is the mechanism by which the user of a name succeeds in referring with the name to its bearer? As the example indicates, whatever the mechanism is, it must be something that can relate the use of a name to its bearer even after the bearer has ceased to exist.

*In modern philosophy of language there are two main views about the nature of the mechanism. On one account, which originated with Frege, a use of a name expresses a conception or way of thinking of an object, and the name refers to whatever object fits, or best fits, that conception or way of thinking. Thus with 'Cassius', for example, I may associate the conception 'the conspirator whom Caesar suspected because of his size' (recalling a famous speech in Shakespeare's *Julius Caesar*). Conception theories are usually called 'sense' theories, after Frege's term 'Sinn'. The other account is the 'historical chain' theory, due to Kripke and Geach. In Geach's words, 'for the use of a word as a proper name there must in the first instance be someone acquainted with the object named.... But...the use of a given name for a given object...can be handed on from one generation to another...Plato knew Socrates, and Aristotle knew Plato, and Theophrastus knew Aristotle, and so on in apostolic succession down to our own times. That is why we can legitimately use "Socrates" as a name the way we do' (1969-70: 288-9).*

1 Sense theories: Introduction

The idea that names express reference-determining conceptions or 'senses' originated with Frege (1892). But Frege was led to it not by wondering about the mechanism of reference, but rather about the difference in 'cognitive value' between two true identity statements of the respective forms $a = a$ and $a = b$ (for example, 'Clark Kent is Clark Kent' and 'Clark Kent is Superman'). Statements of the former sort, he says, are a priori, while the latter 'often contain very valuable extensions of our knowledge and cannot always be established a priori' ([1892] 1980: 56). Frege takes this to show that the propositions expressed by $a = a$ and $a = b$ cannot be the same even when a and b are names of the same person. Yet if the names do corefer, there is no difference in the references of the constituents of $a = a$ and $a = b$; moreover, these two sentences are assembled in the same way. So either they cannot express different propositions, or else - and this is the inference Frege drew - what determines the proposition a sentence S expresses cannot just have to do with the structure of S and the references of its constituent words and phrases (see Frege, G. §§3-4; *Sense and reference* §1).

Frege's proposal was that in addition to possessing a reference, any meaningful expression has a sense, the sense being a 'way of thinking' of the reference. The proposition a sentence expresses is determined by the senses of the words in it, not their references. Thus 'Clark Kent is Clark Kent' and 'Clark Kent is Superman' express different propositions because 'Clark Kent' and 'Superman' express different senses, senses which happen to be ways of thinking of the same reference. Taking the Superman fiction to be fact, the sense of 'Clark Kent' might be 'the mild-mannered reporter on *The Daily Planet* who has a crush on Lois Lane' while the sense of 'Superman' might be 'the blue-suited extraterrestrial who flies'. Here we specify different conceptions of the same individual in two definite descriptions (expressions of the form 'the so-and-so' - see *Descriptions*). The proposition that Superman is Clark Kent therefore has the content that the blue-suited extraterrestrial who flies is the mild-mannered reporter on *The Daily Planet* who has a crush on Lois Lane, which may indeed be, in Frege's phrase, an extension of our knowledge.

It was Frege's view that in an ideal language each name would have a fixed sense and reference for everyone; in an ordinary natural language, there are names that fail to refer, and users may fail to agree on the sense of a name for a specific individual, which Frege thought of as deficiencies of natural language. In a modification of Frege's views, Searle (1958) allowed that a name may be associated with a whole range of descriptions, different users using different ranges; the bearer of the name need not satisfy all the conditions mentioned in the descriptions, only 'a sufficient number', an intentionally vague condition.

Sense theories along such lines as these have been called ‘famous deeds’ sense theories. Famous deeds sense theories appear to have been conclusively refuted by Kripke (1972, 1980). Kripke objects to such sense theories both as they respond to Frege’s own puzzle about the difference between $a = a$ and $a = b$, and as they address our initial question about the mechanism of reference. Let us consider these two issues in turn.

2 Sense theories: Do names express senses?

According to Kripke, the propositions one expresses using names do not involve senses or ways of thinking expressed by the names, otherwise some such propositions would be both metaphysically necessary and a priori, which they are clearly not. For example, suppose the sense of ‘Aristotle’ is ‘the pupil of Plato who tutored Alexander’. Then the sentence

(1) Aristotle was a pupil of Plato

would express the proposition with the content

(2) The pupil of Plato who tutored Alexander was a pupil of Plato.

But (2) is, in a certain sense, necessary, while (1) is not. There is no way things could have gone in which (a) a unique pupil of Plato who tutored Alexander exists and (b) that person was not a pupil of Plato. On the other hand, there are many ways things could have gone in which (a’) Aristotle exists but (b’) Aristotle was not a pupil of Plato (for example, he died young). In other words, granted that there is such a person as Aristotle, he may be a pupil of Plato or he may not be. But granted that there is such a person as the pupil of Plato who tutored Alexander, it follows that he is a pupil of Plato.

This example brings out Kripke’s famous distinction between ‘rigid’ and ‘non-rigid’ designators. In thinking about or describing other ways things could have gone (other possible worlds) we use a proper name such as ‘Aristotle’ consistently to denote the same person; this makes proper names rigid designators. But we use definite descriptions such as ‘the pupil of Plato who tutored Alexander’ differently. With respect to the actual world, this description picks out Aristotle; with respect to a possible world where someone else is the one and only pupil of Plato who tutored Alexander, the description picks out that other person, not Aristotle; and with respect to a possible world where either Plato or Alexander does not exist, the description fails to pick out anyone, even if Aristotle does exist. Hence such descriptions are non-rigid designators. Certain descriptions, such as ‘the positive square root of 9’, are as rigid as proper names, but this is on account of their subject matter, not their semantic role. It is because the typical famous deeds description is non-rigid that the contrast between pairs such as (1) and (2) *vis-à-vis* necessity and contingency arises (see [Reference §2](#)).

Just as (1) and (2) differ in modal status, they differ epistemically: bracketing the question of existence, (2) is, in a limited sense, knowable a priori, while (1) is not. That is, granted that there was such a person as Aristotle, it is a further, empirical question whether he was a pupil of Plato (there may be a controversy among historians about this). But granted that there was such a person as the pupil of Plato who tutored Alexander, it is not a further question, *a fortiori* not an empirical one, whether he was a pupil of Plato.

The same objections arise to Searle’s modified version of the sense theory. For example, no matter what range of descriptions we associate with a name, they will generate statements that are a priori in the manner of (2). If being ϕ logically implies being ϕ' , then

(3) The thing which is F -and- G or G -and- H is F' or G' or H'

is essentially the same as (2), just more complicated. But if the predicates F , G and H encapsulate famous deeds, the corresponding

(4) NN is F' or G' or H' ,

where ‘ NN ’ is the name with which ‘the thing which is F or G or H ’ is associated, will be no more a priori than (1) ([Kripke 1980](#)).

3 Sense theories: Do senses determine reference?

Kripke (1972, 1980) demonstrates another flaw in famous deeds sense theories, namely, that they do not provide an adequate answer to the question about the mechanism of reference. This is because (1) the likely candidate for the sense of a name may pick out an object which is not in fact the name's bearer, or may fail to pick out anything, and (2) we can succeed in referring with a name even when we do not have a 'famous deeds' description associated with it.

Kripke illustrates (1) with two examples. If any description is associated with the name 'Gödel' it is 'the discoverer of the incompleteness of arithmetic'. Does this mean that 'Gödel' refers to that person? What if the theorem was actually proved by Schmidt, who died in strange circumstances, and Gödel got hold of Schmidt's work and represented it as his own? The very fact that we can understand this 'what if' shows that the reference of 'Gödel' is not fixed as whoever discovered the incompleteness of arithmetic. And though this example is fictional, there are similar actual cases: Peano's Axioms are not due to Peano; and Einstein was not the inventor of the atomic bomb (Kripke 1980). For a case where there are descriptions that do not pick out any object although the relevant name still refers, Kripke gives the example of the prophet Jonah, who really existed (according to the scholarly consensus), but whose career as described in the Bible is essentially fictitious. Again, the mere intelligibility of the claim 'Jonah was a historical person but everything uniquely identifying that the Bible says about him is fictitious' is enough to show that the reference of 'Jonah' is not fixed as the Hebrew prophet who was swallowed by a whale, or by any other condition, no matter how complicated or disjunctive, derived from the Book of Jonah.

As for (2), successful reference without associated (definite) descriptions, Kripke points out that most people can use the names 'Richard Feynman' and 'Murray Gell-Mann' to refer to those two people, but that, at best, all the typical person knows about either is that he is a famous physicist who won a Nobel Prize (this was before Feynman achieved popular fame for his role in the *Challenger* disaster inquiry). So we have difference in reference with no difference in associated descriptions; therefore associated descriptions are not at the heart of how proper name reference works.

Perhaps these examples only establish such a conclusion for descriptions that encapsulate famous deeds. Kripke considers some other approaches and concludes that they violate an important non-circularity condition: that candidate descriptions must not themselves embed the notion of reference in a way that cannot eventually be eliminated. For example, we might suggest that the reference of 'Socrates' is fixed as 'the man called "Socrates"'. Since 'called' just means 'referred to as', we do not explain how reference to Socrates is possible in this way: what we want to know is how Socrates *gets to be* the man referred to as 'Socrates'. So this goes nowhere as a proposal about the mechanism of reference, and Kripke plausibly argues that the same would be true for more complicated versions of the idea, for example, that 'Gödel' refers to the person to whom the proof of the incompleteness of arithmetic is commonly attributed. So these attempts at a non-famous-deeds description theory fail. However, in view of Kripke's critique of the famous deeds approach, it seems that if any sense theory is to work, it will have to be a non-famous-deeds account of some sort.

4 Historical chains

The main competing account of the mechanism of reference is the Geach-Kripke historical chain account, in which competence to refer with the name is transmitted across generations in the style adverted to in the quotation from Geach in this entry's prologue (1969-70). The historical chain account is sometimes called the 'causal' chain account, since it is held that the links in the chain are forged by transactions of a causal sort. In Kripke's own version, which he says is a 'picture' rather than a theory (1980: 97), a name is introduced into a community by some 'initial baptism' of an object with the name, and then the name is passed on from link to link, it being required at each step that the receiver of the name 'intend...to use it with the same reference as the [person] from whom he heard it' (1980: 96).

If we regard being told about, or otherwise hearing about, an object, as a way of becoming 'acquainted' with it, then the Geach-Kripke account instantiates an approach championed by Russell (1918), who made acquaintance with an object necessary for referring to it. However, Russell's notion of acquaintance was rather idiosyncratic: apart from my own sense-data, my self, universals and perhaps the present moment, I lack acquaintance with things, according to Russell. Since the ordinary names I use, ostensibly for other people and things, are not names of sense-data, Russell claimed that in a 'logical' sense, ordinary names are not 'proper' names. Rather, he

suggested, they are definite descriptions in disguise (see [Russell, B.A.W. §9](#)). For this reason, Kripke sometimes calls the view he opposes the ‘Frege-Russell theory of names’. The point to bear in mind is that Frege and Russell had different accounts of how reference works (so there is no ‘Frege-Russell theory of reference’), but because on Russell’s account ordinary names do not really refer, the two philosophers end up saying similar-sounding things about such names.

Returning to Kripke’s account of the mechanism of reference in terms of informational exchanges in which the intention to preserve reference is present, the obvious question is whether it is any improvement on circular description theories. After all, the notion of reference enters explicitly into Kripke’s necessary condition for successful passing on of the name, and it also enters at the start of the chain, where some kind of demonstrative reference to the object being baptized is standardly made. To put the same question another way, if this account is explanatory, would there be anything wrong with a description theory which attributed to a name ‘*NN*’ and a user of the name *U* the sense ‘the object at the start of the chain of reference-preserving links by which I came into mastery of this name’, in which ‘*I*’ refers to *U*?

One problem with this version of a description theory is that no ordinary speaker of a natural language associates any such description with a proper name, since the description embodies a philosophical theory, and it is difficult to see how one could justify claiming that the association is ‘implicit’. Here there is a contrast with Kripke’s proposal, which only requires speakers to have the intention to preserve the reference of the new name they have just learned, and it is surely plausible that speakers do have such an intention. Still, there is some sense in which the fundamental nature of the mechanism of reference is left unexplained. For instance, what exactly is it about an actual baptismal service, or about parents announcing their choice of name, that causes a new ‘common currency’ (see [Kaplan 1990](#)) name to be added to the language? Or is the request for further explanation here a demand for a ‘reductive’ account of reference in terms of non-semantic notions (see [Reference §8](#)), something perhaps impossible?

The causal chain picture is not without other difficulties, as Kripke mentions. For instance, straightforward application of the picture could lead to the conclusion that ‘Santa Claus’ is the name of a certain central European king, which seems wrong. Evans ([1973](#)) proposed a different causal account on which the ‘causal source’ of the information associated with the name determines whom it refers to, though in a possibly complex way. Evans observes that the reference of a name in a linguistic community can change over time if it was originally a name for *x* that was mistakenly but consistently misapplied to *y*: *y* would ultimately become the bearer of the name, being the dominant causal source of the information associated with it (unless, as Evans notes, younger members of the community defer in their use of the name towards those for whom it was once a name for *x*). Evans subsequently elaborated his account ([1982: ch. 11](#)).

5 Direct reference and Frege’s puzzle

A striking feature of the Geach-Kripke picture is that it leaves no role for sense to play as that which determines reference. The only candidate for the ‘meaning’ of a name is therefore the name’s reference itself. A theory which claims that the meaning of a name is just its reference, pure and simple, is often called a ‘direct reference’ theory, ‘direct’ signifying that reference is not mediated via sense. Such theories trace their origins through Russell back to Mill ([1843](#)); more recently, the idea is prominent in the writings of Marcus on reference ([1961](#)), and has subsequently been developed and defended in a sustained form by Salmon ([1986](#)) and Soames ([1994](#)).

The main problem which direct reference theories face is the puzzle about the difference between $a = a$ and $a = b$. The general ‘failure of substitutivity’ puzzle is the puzzle of how it is possible for the meanings of two sentences to differ if the sentences have the same structure and their corresponding parts have the same meaning. For example,

(5) It is self-evident to any rational thinker that if Superman exists, then Superman = Superman

seems true and ‘Superman’ and ‘Clark Kent’ have the same reference. Therefore, according to direct reference theory, (5) should have the same meaning as

(6) It is self-evident to any rational thinker that if Superman exists, then Superman = Clark Kent.

But at first sight, (6) does not even have the same truth-value as (5), never mind the same meaning.

Failure of substitutivity is handled straightforwardly on a sense theory of names which allows that coreferential names can have different senses. The claim would be that (not merely the reference but) the sense of 'Superman' enters into the truth-condition of (5), while in (6) the sense of 'Clark Kent' is also involved. There are various mechanisms which might be invoked here, Frege's own being the simplest: in (5) and (6), according to Frege, the proper names refer to their *senses*, rather than to the person they normally refer to (see [Propositional attitude statements §1](#); [Sense and reference §5](#)). But the trouble with any explanation of substitutivity failure for proper names that invokes senses is that sense theories of proper names have been so thoroughly battered by Kripke.

Direct reference theory seems to have less to work with to explain the semantic difference between (5) and (6). On perhaps the best known version of the theory, [Salmon's](#), it is simply denied that there is a semantic difference between (5) and (6). The appearance that substitution changes truth-value in cases such as (5) and (6) is explained as being due to pragmatic effects (see [Pragmatics](#)). Apparently, the only alternative for a direct reference theorist is to identify some assumption in the argument that substitution of coreferring names ought not to change meaning and deny that assumption. This strategy produces a possible target. We suggested that what makes substitutivity failure puzzling is that sentences with the same structure whose corresponding parts have the same meaning should themselves have the same meaning. But this presupposes that the meaning of the entire sentence is wholly determined by the meaning of its parts and their manner of composition (see [Compositionality](#)). However, the notion of structure alluded to in the phrase 'manner of composition' is one which is sensitive only to syntactic categories of expression, not to the identity of expressions. There is another, more 'logical' notion of structure, on which use of a different word, even one with the same meaning, can disrupt structure ([Putnam 1954](#)). In this logical sense, the structure of 'if Superman exists, then Superman = Superman' is 'if *Et* then *t t*', while 'if Superman exists, then Superman = Clark Kent' has the different structure 'if *Et* then *t = t*Cast;'. If we hold that substitution is acceptable only when it does not change logical structure, then from (5) we can infer merely

(7) It is self-evident to any rational thinker that if Clark Kent exists, then Clark Kent = Clark Kent

which is presumably true if (5) is. Of course, this does not help with cases of substitutivity failure in which there is only one occurrence of the name being substituted ('Lois believes that Superman is an extraterrestrial'), but some direct reference theorists have tried to develop a more general notion of structure for these cases ([Taschek 1995](#)).

It is the handling of substitutivity puzzles that is the deciding issue between sense theories and direct reference theories of proper names. Sense theorists need to find a viable account of the senses of names and direct reference theorists need to find a persuasive account either of why substitutivity fails in cases such as (5), or else of why it gives such a convincing appearance of doing so.

See also: [De re/de dicto](#); Kripke, S.A.; [Logical and mathematical terms, glossary of](#)

GRAEME FORBES

References and further reading

- Burge, T.** (1973) 'Reference and Proper Names', *Journal of Philosophy* 70: 425-39. (Argues that, semantically, proper names should be understood as certain kinds of predicate.)
- Deutsch, H.** (1989) 'On Direct Reference', in J. Almog, J. Perry and H.K. Wettstein (eds) *Themes from Kaplan*, Oxford and New York: Oxford University Press, 167-95. (Applies a Kaplan-style approach to context-dependence to proper names.)
- Devitt, M.** (1981) *Designation*, New York: Columbia University Press. (Develops a causal theory of reference using the 'language of thought' hypothesis.)
- Evans, G.** (1973) 'The Causal Theory of Names', *Proceedings of the Aristotelian Society*, supplementary vol. 47: 187-208; repr. in *Collected Papers*, Oxford: Oxford University Press, 1985, 1-24. (A very accessible and insightful discussion of Kripke's account.)
- Evans, G.** (1979) 'Reference and Contingency', *The Monist* 62: 161-84; repr. in *Collected Papers*, Oxford: Oxford University Press, 1985, 178-213. (Discusses a category of proper name not considered in this article, which Evans calls 'descriptive' names. These are names whose reference is explicitly fixed by a definite description, and which figure in Kripke-style examples of the contingent a priori.)
- Evans, G.** (1982) *The Varieties of Reference*, Oxford and New York: Oxford University Press. (Published after

- Evans' death; some parts are unfinished. As well as Evans' final thoughts on proper names, includes influential discussion on Frege, Russell and the notion of an 'object-dependent' thought.)
- Forbes, G.** (1990) 'The Indispensability of Sinn', *Philosophical Review* 99: 535-63. (Develops a non-famous-deeds sense theory of names and applies it to substitutivity failure.)
- Frege, G.** (1892) 'Über Sinn und Bedeutung', *Zeitschrift für Philosophie und philosophische Kritik* 100: 25-50; trans. M. Black, 'On Sense and Reference', in *Translations from the Philosophical Writings of Gottlob Frege*, ed. P.T. Geach and M. Black, Oxford and New York: Blackwell, 3rd edn, 1980. (Widely regarded as the paper which inaugurated modern philosophy of language.)
- Geach, P.T.** (1969-70) 'The Perils of Pauline', *Review of Metaphysics* 23: 287-300; repr. in *Logic Matters*, Berkeley, CA: University of California Press, 1972, 153-65. (States the historical chain view and discusses the use of names in intentional contexts.)
- Jubien, M.** (1993) *Ontology and the Fallacy of Reference*, Cambridge and New York: Cambridge University Press. (An unusual but interesting account of reference, based on 'revisionist' metaphysics.)
- Kaplan, D.** (1990) 'Words', *Proceedings of the Aristotelian Society*, supplementary vol. 64: 93-117. (Includes an original and accessible discussion of how names should be individuated.)
- Kripke, S.A.** (1972) 'Naming and Necessity', in D. Davidson and G. Harman (eds) *Semantics of Natural Language*, Dordrecht: Reidel, 252-355. (One of the most influential and widely discussed works in post-war philosophy.)
- Kripke, S.A.** (1979) 'A Puzzle About Belief', in A. Margalit (ed.) *Meaning and Use*, Dordrecht: Reidel. (Develops two famous cases - the London/Londres case and the Paderewski case - in defence of the thesis that substitution puzzles do not necessarily tell against the principle of substitutivity.)
- Kripke, S.A.** (1980) *Naming and Necessity*, Oxford and New York: Blackwell. (The book version of Kripke (1972).)
- Marcus, R.B.** (1961) 'Modalities and Intensional Languages', *Synthese* 13: 303-22. (Includes an early statement of contemporary direct reference theory.)
- Mill, J.S.** (1843) *A System of Logic*, London: Parker; repr. in *The Collected Works of John Stuart Mill*, ed. J.M. Robson, London: Routledge, 1996, vol. 7. (Draws a famous distinction between the denotation and the connotation of a word and defends a metalinguistic analysis of identity sentences.)
- Putnam, H.** (1954) 'Synonymy and the Analysis of Belief Sentences', *Analysis* 14: 114-22. (Attacks compositionality assumptions in substitution puzzles.)
- Recanati, F.** (1993) *Direct Reference*, Oxford and New York: Blackwell. (Distinguishes a sense in which names are directly referential from two other semantically relevant senses in which they express modes of presentation. Also includes a good discussion of how 'what is said' by an utterance is determined.)
- Russell, B.A.W.** (1918) 'The Philosophy of Logical Atomism', *The Monist* 28: 495-527; 29: 32-63, 190-222, 345-80; repr. in *Logic and Knowledge: Essays 1901-1950*, ed. R.C. Marsh, London: Routledge, 1992. (One of many possible sources of Russell's views, which changed throughout his long life. Lectures II and VI are the most germane to this entry.)
- Salmon, N.** (1986) *Frege's Puzzle*, Cambridge, MA: MIT Press. (The most complete attempt by a direct reference theorist to account for the problem of (apparent) failure of substitutivity.)
- Searle, J.** (1958) 'Proper Names', *Mind* 67: 166-73. (A significant elaboration of broadly Fregean themes.)
- Soames, S.** (1994) 'Attitudes and Anaphora', in J.E. Tomberlin (ed.) *Philosophical Perspectives: 8 Logic and Language*, Atascadero, CA: Ridgeview, 251-72. (Within the context of direct reference theory, investigates the behaviour of anaphoric pronouns in propositional attitude ascriptions. Mostly nontechnical, but rigorous.)
- Strawson, P.F.** (1959) *Individuals*, London: Methuen. (Chapter 6 includes an elaborated Fregean account that Kripke often uses as a stalking horse.)
- Taschek, W.** (1995) 'Belief, Substitution, and Logical Structure', *Noûs* 29: 71-95. (Extends the approach of Putnam (1954) to account for a wider range of substitution puzzles.)

Property theory

Traditionally, a property theory is a theory of abstract entities that can be predicated of things. A theory of properties in this sense is a theory of predication - just as a theory of classes or sets is a theory of membership. In a formal theory of predication, properties are taken to correspond to some (or all) one-place predicate expressions. In addition to properties, it is usually assumed that there are n -ary relations that correspond to some (or all) n -place predicate expressions (for $n \geq 2$). A theory of properties is then also a theory of relations.

In this entry we shall use the traditional labels 'realism' and 'conceptualism' as a convenient way to classify theories. In natural realism, where properties and relations are the physical, or natural, causal structures involved in the laws of nature, properties and relations correspond to only some predicate expressions, whereas in logical realism properties and relations are generally assumed to correspond to all predicate expressions.

Not all theories of predication take properties and relations to be the universals that predicates stand for in their role as predicates. The universals of conceptualism, for example, are unsaturated concepts in the sense of cognitive capacities that are exercised (saturated) in thought and speech. Properties and relations in the sense of intensional Platonic objects may still correspond to predicate expressions, as they do in conceptual intensional realism, but only indirectly as the intensional contents of the concepts that predicates stand for in their role as predicates. In that case, instead of properties and relations being what predicates stand for directly, they are what nominalized predicates denote as abstract singular terms. It is in this way that concepts - such as those that the predicate phrases 'is wise', 'is triangular' and 'is identical with' stand for - are distinguished from the properties and relations that are their intensional contents - such as those that are denoted by the abstract singular terms 'wisdom', 'triangularity' and 'identity', respectively. Once properties are represented by abstract singular terms, concepts can be predicated of them, and, in particular, a concept can be predicated of the property that is its intensional content. For example, the concept represented by 'is a property' can be predicated of the property denoted by the abstract noun phrase 'being a property', so that 'Being a property is a property' (or, 'The property of being a property is a property') becomes well-formed. In this way, however, we are confronted with Russell's paradox of (the property of) being a non-self-predicable property, which is the intensional content of the concept represented by 'is a non-self-predicable property'. That is, the property of being a non-self-predicable property both falls and does not fall under the concept of being a non-self-predicable property (and therefore both falls and does not fall under the concept of being self-predicable).

1 Logical realism and conceptualism in second-order logic

Frege (1879) was among the first to formalize a theory of properties and relations. His theory amounts to a version of standard second-order predicate logic (see **Second- and higher-order logics**). By 'standard' here is meant an axiomatic formulation that contains among its valid theses all instances of the full impredicative comprehension principle (CP), in which bound predicate variables may occur, so that a property or relation can be specified in terms of a totality to which it belongs. Formally, CP is described as follows:

$$(\exists F)(\forall x_1) \dots (\forall x_n)[F(x_1, \dots, x_n) \leftrightarrow \phi],$$

where F is an n -place predicate variable not occurring free in ϕ and $x_1, \dots, x_n$ are distinct individual variables.

By CP, every open formula ϕ containing n free individual variables is said to represent a property or relation, which, using the λ -abstraction operator (see **Lambda calculus**), may also be represented by a λ -abstract, $[\lambda x_1 \dots x_n \phi]$, as a complex n -place predicate expression. (Here λ is an operator that binds n individual variables $x_1, \dots, x_n$, and when affixed to a formula ϕ results in a complex predicate expression in which all of the occurrences of $x_1, \dots, x_n$ in ϕ are bound. For example, the complex predicate of being objects x and y such that $F(x, y) \rightarrow (\exists G)[G(x, y) \wedge G(y, x)]$

can be represented by the λ -abstract

$$\{\lambda xy(F(x, y) \rightarrow (\exists G)[G(x, y) \wedge G(y, x)])\}.$$

The λ -abstract is taken as a valid substituent of the bound n -place predicate variables in universal instantiation

and to satisfy the principle of λ -conversion,

$$(\forall x_1) \dots (\forall x_n)([\lambda x_1 \dots x_n \phi](x_1, \dots, x_n) \leftrightarrow \phi).$$

Because every open formula constructed from atomic formulas in terms of the logical connectives - including quantifiers for individual and predicate variables - is assumed to represent a property or relation in such a theory, the theory is said to amount to a version of 'logical realism' (regarding the existence of universals).

Not every property theory or formal theory of predication will validate CP; that is, not every open formula need represent a property or relation in such a theory. In standard *predicative* second-order logic, for example, formulas containing bound predicate variables are not allowed to occur in the comprehension principle - a constraint that is supposed to restrict the assumption of properties to those that can be represented by first-order formulas where predicate quantifiers do not occur. No λ -abstract in which predicate quantifiers occur will be well-formed in such a theory, and the principle of universal instantiation for predicate quantifiers will be restricted to formulas that contain no bound predicate variables. Formulas with predicate quantifiers can be used to specify properties and relations in *ramified* predicative second-order logic - but only as values of predicate variables of a higher 'level' in the ramified hierarchy, so that no property or relation can be specified in terms of a totality encompassing that property or relation. (See Church 1956: §58.) The motivation for such a restriction is unclear, however, if properties and relations are assumed to exist independently of language and thought. In fact, a substitutional interpretation of predicate quantifiers, with predicate variables having only first-order formulas as their substituends, can be given for standard predicative second-order logic - and a similar related interpretation can be given for its ramification - so that the restriction is equivalent to a nominalistic theory of predication, according to which there are no universals - that is, no properties or relations - beyond the predicate expressions of first-order languages. (See Cocchiarella 1986: ch. 1; 1989: §3.)

In a constructive conceptualist theory of predication, where concepts are understood as the rule-following cognitive capacities underlying the use of predicate expressions, the intent of such a restriction on predicate quantifiers is to preclude the possibility of forming impredicative concepts, that is, concepts that presuppose or conceptually involve a totality to which they belong (such as the concept of a least upper bound of a class of real numbers that has upper bounds). The predicative logic of 'constructive conceptualism' can be formulated as a nonstandard predicative second-order logic in which not all predicate expressions - now including all λ -abstracts, even those in which bound predicate variables occur - are assumed to represent a predicative concept. Such a logic is 'free of existential presuppositions' regarding predicate expressions; and, in particular, no presumption is made independently of meaning postulates as to which predicate constants stand for predicative concepts and which do not. Such a predicative logic can also be ramified, so that concepts can be formed in terms of quantification over concepts formed at a prior 'stage' of the ramified hierarchy; but each 'stage' of the ramification differs from the 'levels' of ramification in standard ramified second-order logic in being 'free of existential presuppositions' regarding predicate expressions. Such a nonstandard predicative logic can also be seen as part of a pattern of conceptual development in which impredicative concept-formation becomes possible, that is, as leading to the validation of CP in a 'holistic' conceptualism - if, as seems warranted given the history of mathematics, a pattern of reflective abstraction in which an idealized transition to a limit is acknowledged to be conceptually possible. (See Cocchiarella 1986: ch. 2; 1989: §§7-8.)

The nonstandard predicative logic of constructive conceptualism, and, similarly, the impredicative logic of holistic conceptualism, will contain a property theory if it is assumed - in a framework of conceptual intensional realism, for example - that properties and relations are intensional objects and that most (if not all) concepts will have such an intensional object as their intensional content. As intensional objects, properties and relations will then be constituents of propositions as the intensional contents of sentences. But, as already noted, as intensional objects, properties and relations will be the objects that nominalized predicates denote as abstract singular terms (a topic we will return to in §4 below).

2 Natural realism

There is another conception of properties and relations for which it is inappropriate to construe them as intensional objects, and in particular as constituents of propositions. On this alternative, properties and relations are components of states of affairs (which are part of the natural order, as opposed to propositions, which are part of

the intensional order), and in that regard they are the physical, or natural, properties and relations involved in the laws of nature. Such physical or natural properties and relations may correspond to some of our concepts or predicate expressions, but in general not every concept or predicate expression will even purport to represent, still less have, such a natural property or relation corresponding to it. No comprehension principle will be valid when the predicate quantifier is taken to refer to a natural property or relation, accordingly, and any assumption that a formula represents a natural property or relation will always amount at best to an empirical hypothesis of an applied scientific theory.

Traditionally (as, for example, in Aristotle's theory), natural properties were said to exist only *in re*, that is, only in the concrete objects that have such properties, which means that

$$(\forall^N F)(\exists x_1) \dots (\exists x_k) F(x_1, \dots, x_k),$$

where ' $\forall^N$ ' is a universal quantifier that refers to natural properties and relations when affixed to predicate variables, is valid in such a theory. For this reason, the theory may be called a 'moderate realism'. This thesis is too strong, however, if such a theory is to allow for physical or natural properties and relations - such as those of certain transuranic elements - that *could* be realized in our world as a matter of a natural or causal possibility, even though in fact they never are. In a *modal* moderate realism, the thesis is weakened to

$$(\forall^N F) \Diamond^c (\exists^e x_1) \dots (\exists^e x_k) F(x_1, \dots, x_k),$$

where ' $\Diamond^c$ ' is a modal operator for a causal or natural possibility, and ' $\exists^e$ ' is an existential quantifier that refers only to existing (concrete) objects when affixed to individual variables - and where individual variables and singular terms are otherwise 'free of such existential presuppositions'. The 'concept' of (concrete) existence, defined as

$$E!(x) =_{\text{df}} (\exists^e y)(x = y),$$

does not represent a natural property in such a theory - though an abstract property in the sense of an intensional object may be assumed as the intensional content of such a concept in conceptual intensional realism (which can be consistently combined with a natural realism in the general framework of 'conceptual realism'). (See [Free logics](#) for more on the logic of singular terms that are 'free of existential presuppositions'.)

3 Actualism and possibilism in modal logical realism

Existence is a property in 'modal logical realism', which is based on a logical or metaphysical notion of possibility and necessity (represented by ' $\Diamond$ ' and ' $\Box$ '; see [Modal logic](#)), and which can be either actualist or possibilist. In *possibilist* modal logical realism, individual variables bound by ' $\forall$ ' or ' $\exists$ ' have possible, as well as actual, objects as their values (and abstract objects as well if such are assumed), and the statement that 'Some things do not exist', formulated as $(\exists x)\neg E!(x)$, is well-formed and meaningful. Predicate variables bound by these quantifiers refer to properties that might apply to nonexisting, as well as existing, objects - though some properties, such as $[\lambda x\neg E(x)]$, apply only to nonexisting objects, whereas others, such as $[\lambda x E(x)]$, apply only to existing (concrete) objects.

In *actualism*, only the actualist quantifiers ' $\forall^e$ ' and ' $\exists^e$ ' are allowed, so that individual variables bound by these quantifiers have only actual (existing, concrete) objects as their values, and predicate variables bound by these quantifiers have only 'existence-entailing' (e-)properties and (e-)relations as their values - and hence

$$(\forall^e F) \Box [F(x_1, \dots, x_n) \rightarrow E!(x_1) \wedge \dots \wedge E!(x_n)]$$

is valid in actualism. Unlike the situation in possibilism, not every open formula or λ -abstract represents an e-property or e-relation, which means that the comprehension principle for actualist modal logical realism must be suitably restricted. (See [Cocchiarella 1989: §10](#).)

4 Nominalized predicates in property theories

In addition to the question of which open formulas or λ -abstracts can be said to represent properties or relations (and can be valid substituends of quantified predicate variables), property theories also differ on the question of whether or not properties and relations have an individual (or 'objectual') as well as a predicable nature, and in particular on whether or not they can also have properties and be 'logical subjects' of predication the way objects

in general are. In Frege's theory, for example, properties and relations are functions from objects to truth-values, where all functions, according to Frege, have an 'unsaturated' nature - a nature that categorially precludes them from being objects, all of which have a saturated nature. Even if predication is not explained in terms of functionality, however, properties and relations can still be construed as having an unsaturated nature that precludes them from being objects. In Wittgenstein's version of 'logical atomism' (see [Wittgenstein 1922](#)), for example, and in natural realism as well, material or natural properties and relations are the nexuses (or modes of configuration) of states of affairs, and as such they cannot also be objects configured in states of affairs (see [Wittgenstein, L.; Logical atomism §1](#)). Similarly, in conceptualism, predicable concepts, as rule-following cognitive structures underlying the use of predicate expressions, are unsaturated cognitive capacities whose exercise (or saturation) in thought and speech is what informs our speech and mental acts with a predicable nature - which means that concepts have only a predicable and not also an individual (or 'objectual') nature. This does not preclude the intensional content of concepts - which are also called properties and relations in the sense of the Platonist, rather than the Aristotelian, tradition - from being objects; namely, the abstract objects that nominalized predicates denote as abstract singular terms, as opposed to the unsaturated concepts that predicates stand for in their role as predicates. Frege held a view something like this in his version of logicism - except that, because of his commitment to an extensional logic, Frege took the objects denoted by abstract singular terms to be classes rather than intensional objects. The important point in all of these views is that even though the universals that predicate expressions stand for in a given formal theory of predication are not themselves objects (values of the individual variables), some (or all) of them may nevertheless have objects correlated with them as the denotata of the same predicate expressions nominalized as abstract singular terms.

The usual way of representing the nominalization of a simple or complex predicate expression as an abstract singular term is by simply deleting the parentheses and commas that are part of that expression's role as a predicate. Thus, whereas $F(x)$ and $R(x, y)$ are formulas in which F and R occur only as predicates, $G(F)$ and $G(R)$ are formulas in which they occur only as singular terms; and in $F(F)$ and $R(F, R)$, they occur first as predicates and then as singular terms. Similarly, where ϕ is a formula, $[\lambda x \phi]([\lambda x \phi])$ is a formula in which $[\lambda x \phi]$ occurs first as a predicate and then as a singular term. A stronger comprehension principle (CP^*_λ) can now be formulated in terms of identity,

$$(\exists F)([\lambda x_1 \dots x_n \phi] = F),$$

from which, by Leibniz's law (applied now even to abstract singular terms), CP is a consequence.

Bertrand Russell, like Frege, also maintained a version of logical realism; but, unlike Frege, Russell rejected the idea that properties and relations have an unsaturated nature that precludes them from being objects (even though he explained the unity of a proposition in terms of a relation occurring only as a relation and not also as an object). Unlike the situation for Frege, in other words, for Russell a nominalized predicate denotes the same property or relation as an abstract singular term that the predicate otherwise stands for in its role as a predicate. This difference is unimportant as far as Russell's paradox of predication is concerned, however, because the paradox, which can be derived from the following instance of CP (and hence from CP^*_λ) in which a nominalized predicate occurs as abstract singular terms,

$$(\exists F)(\forall x)[F(x) \leftrightarrow (\exists G)(x = G \wedge \neg G(x))],$$

applies to both Frege's and Russell's views of predicates and their nominalizations. Russell's way out of his paradox was the ramified theory of types, which involves grammatically categorizing predicate expressions not only in terms of the hierarchy of 'levels' of ramified second-order logic but also in terms of a 'vertical' hierarchy of orders in which predicates that take nominalized predicates as singular terms must always be of a higher order than the latter (see [Theory of types](#)).

Later authors, such as L. Chwistek and F. Ramsey, noted that ramification into levels (of the 'horizontal' hierarchy) is not needed to avoid logical paradoxes such as Russell's, and that a simple theory of types suffices instead - though others, such as A. Church, maintain that the ramified, 'horizontal' hierarchy is still needed for certain intensional or semantic paradoxes. R. Montague (1974) developed an intensional logic for natural language based on a version of simple type theory that was originally formulated by Church. Montague extended Church's theory by adding to it intension- and extension-forming operators (patterned after Frege's *Sinn/Bedeutung*

distinction; see Frege, G. §3), ‘ $\wedge$ ’ and ‘ $\vee$ ’, so that where ξ is a well-formed expression of the theory, $^{\wedge}\xi$ stands for the intension (or sense) of ξ , and $^{\vee}\xi$ stands for its extension (and hence $^{\vee}\wedge^{\vee}\xi = \xi$ is valid). Properties and relations of entities of types $t_1, \dots, t_n$ are then taken as the intensions (or senses) of the classes of n -tuples of those types, and the predication of a property or relation (in intension) is defined as membership of the class that the property or relation has as its extension.

5 Conceptual intensional realism

There are other ways to avoid Russell’s paradox besides type theory. One way is to return to standard second-order logic with nominalized predicates as abstract singular terms and restrict λ -abstracts so that only those that can be homogeneously (h-)stratified from an external (metalinguistic) point of view are well-formed. (An expression is ‘h-stratified’ if it is possible to assign natural numbers to the predicates and singular terms occurring in it so that singular terms flanking the identity sign, as well as all of the subject expressions of any predicate expression occurring in it, are assigned the same number, with the predicate itself assigned the successor of that number.) This amounts to restricting the comprehension principle CP_{λ}^* so that even though the above instance of CP for Russell’s paradox is well-formed, it is not provable, because the λ -abstract

$$[\lambda x(\exists G)(x = G \wedge \neg G(x))]$$

is not h-stratified. The resulting system is equiconsistent with the simple theory of types.

Another way to avoid Russell’s paradox in the context of second-order logic with nominalized predicates is to allow all λ -abstracts, h-stratified or not, to be well-formed and to retain the full comprehension principle CP_{λ}^* , but allow the first-order part of the logic to be free of existential presuppositions regarding singular terms. The principle of λ -conversion must then be qualified as follows,

$$[\lambda x_1 \dots x_n \phi](a_1, \dots, a_n) \leftrightarrow (\exists x_1) \dots (\exists x_n)(a_1 = x_1 \wedge \dots \wedge a_n = x_n \wedge \phi),$$

where no x_i is free in any a_j , for all i, j such that $1 \leq i \leq j \leq n$. Then, even though

$$(\exists F)([\lambda x(\exists G)(x = G \wedge \neg G(x))] = F)$$

is provable as an instance of CP_{λ}^* , that is, even though $[\lambda x(\exists G)(x = G \wedge \neg G(x))]$ stands for a predicable concept in its role as a predicate, all that follows by the argument for Russell’s paradox is that it cannot also denote an object in its role as an abstract singular term, because

$$\neg(\exists y)([\lambda x(\exists G)(x = G \wedge \neg G(x))] = y)$$

is also provable. In conceptual intensional realism, this result means that the attempt to ‘object-ify’ the content (in the sense at least of the truth-conditions) determined by the predicable concept represented by $[\lambda x(\exists G)(x = G \wedge \neg G(x))]$ as an intensional object (or property in the Platonist tradition) must fail. In a reconstruction of Frege’s extensional theory, it means that there is no extension (class) corresponding to the property in question. This is not possible in a reconstruction of Russell’s view, however, because for Russell every nominalized predicate denotes the same property that the predicate stands for. (See Russell 1903: §49.)

Finally, there are also strictly first-order property theories. Some are developed along the lines of Zermelo-Fraenkel set theory, with membership replaced by exemplification as a predication relation, and with the deletion at least of the axiom of extensionality (see Schock 1969), and sometimes also of the axiom of foundation (see Jubien 1989). (See Set theory, different systems of.) Some authors prefer to defer assuming specific axioms for predication and simply extend first-order logic to include nominalized predicates as abstract singular terms, leaving the choice open of how predication might be axiomatically developed as a relation regarding the denotata of such singular terms. (See Bealer and Mönich 1989.) Others develop a first-order theory of properties based on assuming a similarity between revisionist theories of truth and predication as a first-order relation (see Turner 1987).

See also: Logical and mathematical terms, glossary of

NINO B. COCCHIARELLA

References and further reading

- Bealer, G.** (1982) *Quality and Concept*, Oxford: Clarendon Press. (Accessible to readers with an intermediate-level knowledge of logic; presents a theory of properties, relations and propositions as a first-order intensional logic.)
- Bealer, G. and Mönnich, U.** (1989) 'Property Theories', in D. Gabbay and F. Guenther (eds) *Handbook of Philosophical Logic*, Dordrecht: Reidel, vol. 4, 133-251, 1994. (Argues for a first-order theory of properties, relations and propositions in which nominalized forms of complex predicates are included among the singular terms as 'intensional abstracts'.)
- Chierchia, G. and Turner, R.** (1988) 'Semantics and Property Theory', *Linguistics and Philosophy* 11: 261-302. (Develops a multi-sorted first-order theory of properties in which predication is analysed along the lines of a revisionist theory of truth.)
- Chierchia, G., Partee, B. and Turner, R.** (1989) *Properties, Types and Meaning*, vol. 1, Dordrecht: Kluwer. (This collection of papers is the result of a conference held in Amherst in 1986.)
- Church, A.** (1956) *Introduction to Mathematical Logic*, Princeton, NJ: Princeton University Press. (An excellent text for the intermediate student; includes axiomatic and semantic formulations not only of propositional and first-order logic, but also of second-order logic, standard predicative and ramified second-order logic.)
- Cocchiarella, N.B.** (1986) *Logical Investigations of Predication Theory and the Problem of Universals*, Naples: Bibliopolis. (Formal theories of predication are semantically associated in this book with theories of universals for nominalism, logical and natural realism, and constructive and holistic conceptualism. The first part of this book, which can be read by students with an intermediate-level knowledge of logic, deals with predication theories as second-order logics; the second part extends those theories to second-order logic with nominalized predicates as abstract singular terms.)
- Cocchiarella, N.B.** (1989) 'Philosophical Perspectives on Formal Theories of Predication', in D. Gabbay and F. Guenther (eds) *Handbook of Philosophical Logic*, vol. 4, Dordrecht: Reidel, 254-326. (Discusses in philosophical terms some of the formal theories of predication described in part 1 of Cocchiarella (1986).)
- Frege, G.** (1879) *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*, Halle: Nebert; trans. 'Begriffsschrift, a Formula Language, Modelled Upon That of Arithmetic, for Pure Thought', in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967, 1-82. (Accessible to most students, although Frege's two-dimensional notation can be somewhat difficult at first; now recognized as a classic.)
- Frege, G.** (1952) *Translations from the Philosophical Writings of Gottlob Frege*, ed. P.T. Geach and M. Black, Oxford: Blackwell, 3rd edn, 1980. (The essays 'On Concept and Object' and 'Function and Concept' present a good exposition of Frege's view of properties (which he called 'concepts') and relations as unsaturated functions from objects to truth-values; widely accessible.)
- Jubien, M.** (1989) 'On Properties and Property Theory', in G. Chierchia, B. Partee and R. Turner (eds) *Properties, Types and Meaning*, Dordrecht: Kluwer, vol. 1, 159-75. (A development of property theory as first-order theory based on a modified Zermelo-Fraenkel set theory.)
- Montague, R.** (1974) *Formal Philosophy: Selected Papers of Richard Montague*, ed. R.H. Thomason, New Haven, CT: Yale University Press. (See especially the essays 'Universal Grammar' and 'The Proper Treatment of Quantification in Ordinary English' for a development of Montague's intensional logic along the Church-Frege lines. 'Pragmatics and Intensional Logic' includes a formulation of intensional logic along more traditional Russellian lines as a second-order modal logic.)
- Parsons, C.** (1971) 'A Plea for Substitutional Quantification', *Journal of Philosophy* 68 (8): 231-7. (Argues for a substitutional - as opposed to a referential - interpretation of predicate quantifiers for standard predicative second-order logic. Instead of having properties and relations as their values, bound predicate variables have only open formulas with no predicate quantifiers as their substituends.)
- Parsons, T.** (1977) 'Type Theory and Ordinary Language', in S. Davis and M. Mithun (eds) *Linguistics, Philosophy, and Montague Grammar*, Austin, TX: University of Texas Press, 127-51. (This paper illustrates what English would be like if it obeyed the rules of a theory of types.)
- Russell, B.A.W.** (1903) *The Principles of Mathematics*, Cambridge: Cambridge University Press; 2nd edn, London: Allen & Unwin, 1937; repr. London: Routledge, 1992. (Includes an account of Russell's early view of properties, which Russell associated, but did not strictly identify, with propositional functions.)
- Schock, R.** (1969) *New Foundations for Concept Theory*, Uppsala: Library of Theoria, no. 12. (A first-order theory of properties as intensional objects.)
- Turner, R.** (1987) 'A Theory of Properties', *Journal of Symbolic Logic* 52: 455-72. (A first-order theory of properties based on an assumed similarity between predication and revisionist theories of truth.)

Wittgenstein, L.J.J. (1922) *Tractatus Logico-Philosophicus*, trans. D.F. Pears and B.F. McGuinness, London: Routledge, 1961. (The properties and relations of this classic of logical atomism are the nexuses, or modes of configuration, of atomic states of affairs, and as such cannot themselves be configured as objects in states of affairs.)

Propositional attitude statements

Propositional attitude statements - statements about our beliefs, desires, hopes and fears - exhibit certain logical peculiarities. For example, in apparent violation of Leibniz's law of the indiscernibility of identicals, we cannot freely substitute expressions which designate the same object within such statements. According to Leibniz's law, every instance of the following scheme is valid:

$a = b$

$F(a)$

Therefore, $F(b)$

The validity of Leibniz's law seems beyond question. It says, in effect, that if an object has a certain property, then anything identical to that object also has that property. Valid instances abound. But consider the following apparently invalid instance:

(1) *Hesperus is Phosphorus*

(2) *Hammurabi believed that Hesperus often rose in the evening*

(3) *Therefore, Hammurabi believed that 'Phosphorus' often rose in the evening.*

If we take 'Hammurabi believed that...often rose in the evening' to serve as the predicate F and 'Hesperus' and 'Phosphorus' to be a and b respectively, this argument appears to be an instance of Leibniz's law. Yet (3) apparently fails to follow from (1) and (2). Hammurabi believed that Hesperus and Phosphorus were two heavenly bodies not one. And he believed that Hesperus did, but that Phosphorus did not rise in the evening.

We have derived a false conclusion from true premises and an apparently valid law. If that law is really valid, then our argument had better not be a genuine instance of the law. The tempting conclusion, widely accepted, is that we were wrong to construe propositional attitude statements as simple predications. We should not, that is, construe 'Hammurabi believed that...often rose in the evening' to be just a long predicate with the semantic function of attributing some property to the object commonly denoted by 'Hesperus' and 'Phosphorus'. But then the question arises: if attitude reports are not simple predications, what are they? Philosophers have disagreed sharply in their answers. Moreover, their disagreements are intimately connected to a wide range of deep issues about the nature of meaning and reference.

1 Frege on attitude statements

Gottlob **Frege** was among the first to offer systematic explanations of the form and content of propositional attitude statements (**Frege 1892**). Central to Frege's treatment of attitude statements is his distinction between sense and reference (see **Sense and reference**). Frege held that each significant linguistic expression plays two distinct but related semantic roles: it 'denotes' a reference and 'expresses' a sense. For example, the reference of a singular term like a proper name is typically some individual object. A sense, on the other hand, is a way of being given a reference. The sense of a proper name, for example, contains a condition, the unique satisfaction of which by an object o is necessary and sufficient to determine o as its reference. Perhaps the following captures the sense of 'Phosphorus':

o is Phosphorus just in case o is the last celestial object visible in the early morning sky just before sunrise.

The distinction between sense and reference holds not just for proper names but also for sentences. Frege took the reference of a complete declarative sentence to be its truth-value. A declarative sentence is thus a name of sorts - a name for a truth-value. So, for example, the reference of:

(4) *Phosphorus can often be seen in the evening sky*

is the True, as Frege called it. On the other hand, the sense of a complete declarative sentence is supposed to be a thought, according to Frege. The sense of (4) is the thought that Phosphorus can often be seen in the evening sky.

Frege held that that two expressions may denote the same reference while expressing distinct senses. 'Hesperus' and 'Phosphorus' denote the same object, but differ in sense, from Frege's viewpoint. Something like the

following satisfaction condition arguably captures the sense of ‘Hesperus’:

o is Hesperus just in case *o* is the first celestial object visible in the early evening sky after sunset.

Similarly, although sentence (4) and

(5) Hesperus can often be seen in the evening sky

denote the same reference - namely, the True - they express different thoughts. For the thought that Phosphorus can often be seen in the evening sky is, according to Frege, distinct from the thought that Hesperus can often be seen in the evening sky. One indicator of their distinctness is the possibility that a rational cognizer can simultaneously believe the one, while disbelieving the other. Hammurabi was just such a cognizer.

It is worth enumerating three essential characteristics of Fregean thoughts. First, thoughts are the primary bearers of truth-conditions. The very identity of a thought is constituted by its having the truth-conditions that it has. Sentences, by contrast, have their truth-conditions only derivatively, by being associated with a thought, in virtue of the conventions of a language. Second, thoughts are composite structured entities. A thought is a whole, composed of constituent parts. The parts of thoughts are themselves senses. For example, the thought that Hesperus is often visible in the evening has the sense of ‘Hesperus’ as a constituent part. Third, thoughts are the objects of the propositional attitudes. To believe that Hesperus is often visible in the evening is to stand in a certain relation - the believing relation - to the thought that Hesperus is often visible in the evening. To wonder whether Hesperus is often visible in the evening is to stand in a different relation - the wondering relation - to that same thought.

Frege endorses a principle of compositionality for both sense and reference (see [Compositionality](#)). He believes that both the reference and the sense of a complex expression are a function of the references and senses, respectively, of the constituent parts of that sentence (and the way those parts combine to form a whole). Two sentences which differ only by constituents which themselves have the same reference, should have the same truth-value. Similarly, two sentences which differ only by constituents which express the same sense, should themselves express the same thought. Now Frege is convinced that these compositionality principles are nowhere violated. But then the original quandary about attitude sentences remains.

Frege’s way out is to say that despite our initial assessment ‘Hesperus’ as it occurs in (2) and ‘Phosphorus’ as it occurs in (3) do not share a reference. Call the sense and reference of a term as it occurs in direct discourse its ‘customary’ sense and reference. Frege claims that when an expression is embedded within a that-clause, it undergoes a shift in both its sense and its reference. In particular, he holds that obliquely occurring expressions, as he called them, denote not their customary referents, but their customary senses. Moreover, an obliquely occurring term denotes a new sense: what he calls its ‘oblique’ or ‘indirect’ sense. An indirect sense is a mode of presentation not of a term’s customary reference, but of its customary sense. Thus in (2) ‘Hesperus’ refers not to the planet Venus but to the customary sense of ‘Hesperus’. Similarly in (3), ‘Phosphorus’ denotes the customary sense of ‘Phosphorus’. Indeed, a whole sentence undergoes a shift in sense and reference when it is joined with ‘that’ to form a clause. The that-clause in (2), for example, does not denote a truth-value, but the thought customarily expressed by ‘Hesperus often rose in the evening’ when it stands on its own. It denotes, that is, the thought that Hesperus often rose in the evening. And (2) as a whole says that Hammurabi stands in the believing relation to that thought. Similarly, the that-clause in (3) denotes the thought that Phosphorus often rose in the evening. And (3) attributes to Hammurabi the property of believing that thought.

Clearly, if Frege is correct, the invalidity of our original argument does not, after all, violate Leibniz’s law. Since the customary sense of ‘Hesperus’ is not identical to the customary sense of ‘Phosphorus’ our original argument turns out not to involve the substitution of co-referring terms.

2 Referential opacity

Philosophers have sometimes questioned the coherence of Fregean senses. W.V. Quine, for example, has argued that senses and related intensional notions are ‘creatures of darkness’ ([Quine 1956](#)). There is no saying, according to Quine, for arbitrary senses *x* and *y*, when *x* and *y* are the same sense and when they are two distinct senses (see [Quine, W.V. §8; Radical translation and radical interpretation §§2-3](#)). There can be no entities, he insists, where

there are no determinate criteria of identity. But if senses go, so too must Frege's account of propositional attitude statements.

Quine endorses Frege's negative conclusion that obliquely occurring terms do not play their customary referential roles. He offers, however, a markedly different diagnosis of that failure. Like Frege, Quine believes that where reference happens substitutivity reigns - that is, Leibniz's law holds. But unlike Frege, Quine takes the failure of substitutivity within attitude statements at face value. And he concludes, in effect, that reference cannot be happening within the context of propositional attitude statements or that-clauses generally. That-clauses, he claims, are 'referentially opaque'. Embed a term which elsewhere functions referentially within such a clause and substitution is no longer permissible, not because reference shifts, but because the embedded term is stripped of its referential function. Quine suggests, in fact, that it is a mistake to speak of 'embedding' a referring expression within a that-clause at all. He suggests that an expression like 'believed that Hesperus often rose in the evening' functions rather like a primitive one-place predicate with no 'logically germane' constituent parts. In that case, sentences like (2) are misleadingly spelled. A more revealing spelling would treat this predicate as a single, though very long word somewhat like: 'believed-that-Hesperus-often-rose-in-the-evening'. The presence of the string 'Hesperus' in this very long word is merely an accident of orthography. It would be no more correct to regard 'Hesperus' as a grammatical constituent of this very long word, than it would be to regard the word 'cat' as grammatical constituent of the word 'cattle'.

Referential opacity explains another logical peculiarity of that-clauses, according to Quine. Consider the law of existential generalization:

$F(a)$
Therefore, $(\exists x)F(x)$

Informally, this scheme says that if a has F then there is something or other which has F . Existential generalization allows us to infer from the premise that 'John loves Mary' to the conclusion that 'John loves someone or other'. When we try to apply this evidently valid scheme within the context of a propositional attitude verb, we are led into apparent contradiction. Consider the following sentence:

(6) The president suspects that someone is a spy.

As it stands, (6) is subject to two different readings. On one reading, (6) says that the president 'suspects true', as we might put it, a certain statement or proposition - the statement or proposition that spying sometimes happens. This reading has been variously called the '*de dicto*', 'notional', or 'referentially opaque' reading (see *De re/de dicto*). There is a second reading of (6) which attributes to the president not just the generalized suspicion that spying sometimes happens, but the state of being 'suspicious of' some person or other. We might paraphrase this reading of (6) by:

(6b) The president suspects someone of being a spy

or by:

(6c) There is someone whom the president suspects of being a spy.

This reading is variously called the '*de re*', 'relational', or 'referentially transparent' reading. The two readings are not generally equivalent. It is possible to suspect that spying happens even when one has no suspicions about any particular person. The ambiguity in sentences like (6) is sometimes traced to an ambiguity of quantifier scope. On the so-called *de dicto* or opaque reading of (6), the quantifier has narrow scope relative to 'believes'. On the so-called *de re* or transparent reading, the quantifier has wide scope relative to 'believes'. As a first pass, we might represent the *de dicto* reading by (6d) and the *de re* reading by (6e):

(6d) The president suspects that $(\exists x)(x \text{ is a spy})$

(6e) $(\exists x)(\text{The president suspects that } x \text{ is a spy})$.

Quine has argued that constructions like (6e), in which a quantifier sitting outside the scope of the attitude verb 'reaches in' to bind a variable that lies within the scope of that attitude verb, are deeply problematic. We cannot,

he claims, ‘quantify into’ attitude contexts (Quine 1956). Or to put it in slightly different terms, propositional attitude verbs appear to block the interior reach of exterior quantifiers.

Suppose that the president has observed Stanley P. Young, a certain low-level White House aide, behaving in a rather furtive manner. He says with great conviction to the chief of White House security, ‘Stanley P. Young is a spy.’ That seems a sufficient basis for attributing to the president the belief that Stanley P. Young is a spy and thus a sufficient basis for taking this to be true:

(7) The president believes that Stanley P. Young is a spy.

If we construe ‘The president believes...is a spy’ as a complex predicate, then from (7) and the law of Existential Generalization, we should be able to infer:

(8) $(\exists x)(\text{The president believes that } x \text{ is a spy})$.

There may be a number people whom the president believes to be spies. Stanley P. Young is one such person. (8') below would also seem to follow:

(8') $(\exists x)(x = \text{Stanley P. Young} \ \& \ \text{the president believe that } x \text{ is a spy})$.

Now suppose that the president is subsequently introduced to one James Q. Money, a generous contributor to progressive causes and a man of stellar reputation. The president trusts Mr Money. When asked whether Money is a spy, the president denies even the possibility. That denial is grounds for inferring:

(9) The president believes that James Q. Money is not a spy.

From (9) and Existential Generalization we can infer:

(10) $(\exists x)(\text{The president believes that } x \text{ is not a spy})$.

(10') $(\exists x)(x = \text{James Q. Money} \ \& \ \text{the president believe that } x \text{ is not a spy})$.

But unbeknownst to the president, James Q. Money leads a double life. He is none other than Stanley P. Young. Does the president believe or not believe James Q. Money, that is, Stanley P. Young, to be a spy? Intuition pulls in opposite directions. Since James Q. Money just is Stanley P. Young, we should, by Leibniz’s law, be able to substitute ‘Young’ for ‘Money’ in (10') without change of truth-value. Since this substitution happens outside the that-clause, it should be unaffected by the logical peculiarities of such clauses. Yet substitution yields:

(10'') $(\exists x)(x = \text{Stanley P. Young} \ \& \ \text{the president believe that } x \text{ is not a spy})$.

Examples (10'') and (8') are of dubious consistency. No one can both be and not be a spy. So how can anyone consistently both believe a person to be a spy and believe that very person not to be a spy?

The answer lies in referential opacity. Because (7) and (9) are opaque, they do not ascribe to the president incompatible relations to Money, that is, Young. Indeed, opaquely construed, the Quinean insists, neither (7) nor (9) relates the president to Money, that is, Young at all. Recall again the Quinean view that when opaque contexts are perspicuously spelled the illusion that they contain referring expressions as proper constituents will have been dispelled. Indeed, opacity renders quantified sentences like (10') and (10'') utterly nonsensical. For proper respelling reveals that there is no genuine occurrence of an interior variable for the exterior quantifier to bind. An analogy with quotation - which Quine takes to be the referentially opaque context *par excellence* - makes the reason clear. Consider:

(11) ‘x is a spy’ is an open sentence.

Enclosing an expression within quotation marks seals that expression off from the reach of external quantifiers. The quantifier below does not bind the quoted variable:

(12) $(\exists x)(\text{‘}x \text{ is a spy’ is an open sentence})$.

What occurs within the quotation marks is really not a variable *in use*. It is rather a variable *being mentioned* (see

Use/mention distinction and quotation). Similarly, though what occurs in (10') and (10'') bears a certain orthographic resemblance to a variable in use, it is not the real thing. For that reason we can no more quantify into propositional attitude contexts, Quine insists, than we can quantify into the context of quotation.

Quine's approach promises a way out of the threatened contradiction, but only by doing great violence to a number of potent intuitions. It seems frankly incredible that putatively referring terms embedded within a that-clause are mere grammatical illusions. Further, if propositional attitude statements lack logically germane constituents, there will be no accounting for certain systematic commonalities and differences among such statements. Compare and contrast, for example, (7) and (9) above with (13) and (14) below:

(13) The First Lady *doubts that* Young is a spy

(14) The vice-president *wonders whether* Young is a spy.

Example (13) seems to say that the First Lady doubts what the president believes, while (14) says that the vice-president has questions about what the First Lady and the president disagree about. Moreover, the president's belief, the First Lady's assurance and the vice-president's questions seem all to be about one and the same individual. Quine's approach abandons these intuitive judgments.

Quine is well aware of these costs. He believes, however, that we must pay the cost if we are to keep the creatures of darkness - Fregean senses and other intensional entities - at bay (see **Intensional entities**).

3 Innocence regained?

Donald **Davidson** has argued that if we could but regain our pre-Fregean semantic innocence:

it would seem to us plainly incredible that the words 'The earth moves' uttered after the words 'Galileo said that' mean anything different, or refer to anything else, than is their wont when they come in other environments.

(Davidson 1969: 172)

If Davidson is right then we can account for the logical peculiarities of attitude statements without positing Fregean reference shifts. Nor, if he is right, need we allow that otherwise referring terms are stripped of their referential functions when they occur in attitude contexts. Indeed, against Quine, Davidson holds that unless we have already been forced to introduce Fregean senses for other reasons, the peculiarities of attitude contexts provide no additional pressures to posit such entities. So if Davidson is right, Quine's drastic departures are not needed to keep intensions at bay.

Davidson defends what he calls a 'paratactic analysis' of the logical forms of indirect discourse, though his approach can be extended fairly directly to attitude ascriptions. Davidson's central claim is that a sentence like:

(15) Galileo said that the earth moves

is really a parataxis of two sentences. The first sentence of the parataxis in (15) ends with the word 'that'; the second starts with 'the earth'. By adding a bit of punctuation, we can represent (15) in a more perspicuous manner as:

(16) Galileo said that. The earth moves.

Call the first sentence the attribution sentence and the second the content sentence. In the attribution sentence of (16), the 'that' is a demonstrative pronoun which refers (on an occasion) to an utterance of the second sentence of (16).

Sentence (16) will be true, according to Davidson, just in case some utterance of Galileo's makes Galileo and the utterer of (16) 'samesayers'. Suppose that Galileo utters (17) and Smith utters (18):

(17) Eppur si muove.

(18) The earth moves.

In some sense, Galileo and Smith say the same thing. Moreover, if Smith utters (18) and Galileo utters (17) then

‘the earth’ in Smith’s mouth and ‘Eppur’ in Galileo’s mouth have the same reference.

Suppose that Smith wants to attribute Galileo’s statement to Galileo. He can do so simply by saying the same thing as Galileo and then saying that he has just done so - as in the following scenario:

(19) Galileo: Eppur si muove.

(20) Smith: The earth moves. Galileo said that (too).

Here we have Galileo saying, in Italian, that the earth moves. We have Smith saying the same thing in English. And then we have Smith ascribing such a statement to Galileo. He does so by making a statement that turns himself and Galileo into samesayers (to use Davidson’s phrase) and then stating that he has just done so. As matters stand, it remains unclear until Smith utters the second sentence of (20) just what the point of his initial utterance may be. But we can rectify that by reversing the order of the sentences as follows:

(21) Galileo said that. → The earth moves.

The arrow in (21) should be understood as a demonstration accompanying the demonstrative pronoun ‘that’. The demonstrative refers to Smith’s second utterance - the one by which Smith purports to turn himself and Galileo into samesayers. While Smith ‘refers’ to his own utterance, he implicitly ‘quantifies over’ Galileo’s utterances. Smith says, in effect, that the demonstrated utterance of his and ‘some utterance or other’ of Galileo’s make himself and Galileo samesayers.

Davidson’s account yields a relatively straightforward explanation of the invalidity of inferences like the following:

(22) Galileo said that. → The earth moves.

(23) The earth is the third planet from the sun.

(24) Galileo said that. → The third planet from the sun moves.

There is no logical connection between the effect of substitution on the content sentence and the truth-values of the two attribution sentences. As Davidson puts it:

There is no reason to predict, on grounds of form alone, any *particular* effect on the truth of [the attribution sentence] from a change in [the content sentence]. On the other hand, if the [content sentence] had been different in any way at all, [the attribution sentence] *might* have had a different truth-value, for the reference of the ‘that’ would have been changed.

(Davidson 1969: 172)

Though Davidson’s central idea is quite ingenious, there are both syntactic and semantic grounds for scepticism. For example, content clauses are not generally introduced by demonstratives. Neither ‘whether’ in ‘Galileo wondered whether the earth moves’ nor ‘for... to’ in ‘I would prefer for you to leave’ can plausibly be analysed as overt demonstratives. Moreover, demonstratives cannot in general be deleted, but the ‘that’ in ‘... said that...’ can. Compare the following:

(25) Davidson saw that man over there.

(26) Davidson said that the paratactic analysis is true.

(27) ?Davidson saw man over there.

(28) Davidson said the paratactic analysis is true.

It is also difficult to see how to extend the paratactic analysis to *de re* ascriptions. Consider (29):

(29) Every philosopher believes that he is wise.

Example (29) says, in effect, that every philosopher has a high opinion of himself. But if we interpret (29) along Davidsonian lines as:

(30) Every philosopher believes that. He is wise.

we get something rather different.

Finally it is clear that we need to know more about the samesaying relation. If the paratactic analysis is to overcome the need to introduce Fregean senses and the like, then samesaying cannot just be a matter of expressing the same Fregean sense. Davidson owes us an account of samesaying which eschews senses and other intensional entities altogether. Davidson is fully aware of this debt and labours mightily to pay it (Davidson 1969).

4 Structured complexes

A class of theories which has recently gained favour with a number of philosophers is what might be called the structured complex approach (see *Intensional logic* §5). Advocates of such theories maintain that attitude verbs express relations between agents and complex hierarchically structured entities of one sort or another. The core idea dates back to Carnap (1947). It has resurfaced, much modified, in Cresswell (1984), Segal (1989), Richard (1990), Higginbotham (1991) and Larson and Ludlow (1993). Different versions of this approach posit different structured complexes. Carnap posited hierarchically structured intensions. Others pair semantic values with something else: sometimes a sentence, sometimes a description of a sentence, sometimes a structured intension. Larson and Ludlow, for example, hold that that-clauses specify Interpreted Logical Forms (ILFs). An ILF is a kind of product of a phrase structure tree for a sentence and the sequence of semantic values for the semantically valued constituents of that sentence. On their approach, the sentence:

(31) Pierre admires London

will be associated with the ILF in Figure 1:

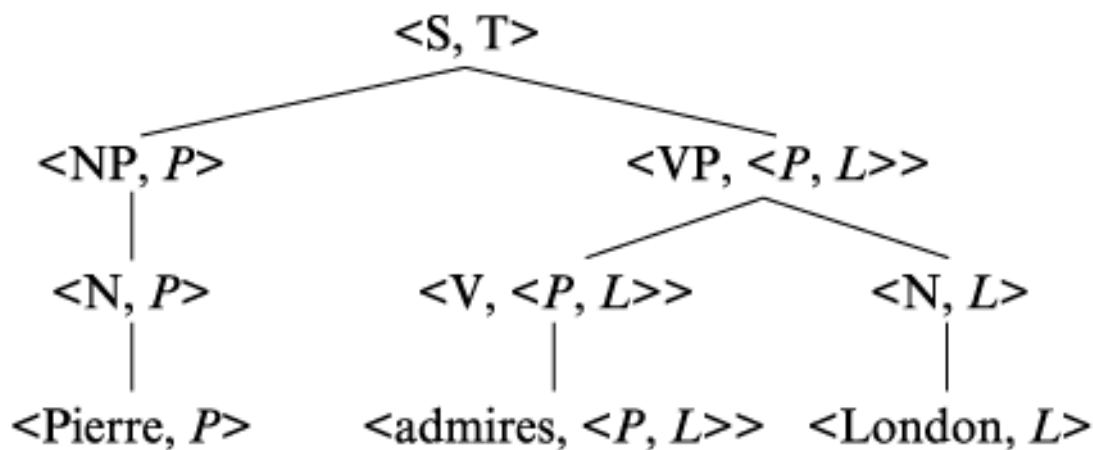


Figure 1

We need not worry about the exact formation rules that generate this ILF. The crucial point is that each node in this tree is labelled by a pair whose left member is either a lexical item (at the bottom-most nodes) or a phrase marker (at all other nodes) and whose right member is a semantic value. For example, the pair $\langle \text{Pierre}, P \rangle$ is the pair whose left member is the word 'Pierre' and whose right member is the reference of 'Pierre' (namely, Pierre himself). The pair $\langle \text{admires}, \langle P, L \rangle \rangle$ has left member 'admires' and a right member consisting of a pair whose members are Pierre and London. Such trees are intended to represent a sequence of semantic values and the order in which those semantic values are associated with the expressions whose values they are. Because lexical items (words themselves), and not just phrase makers, are constituents of ILFs, the following two sentences will have distinct ILFs:

(32) Hesperus rises in the evening.

(33) Phosphorus rises in the evening.

Nonetheless, (32) and (33) have, according to Larson and Ludlow, the same semantic content. That is, (32) and (33) have formally distinct but semantically equivalent ILFs. Armed with the notion of formally distinct but semantically equivalent ILFs, Larson and Ludlow are able to explain how it is possible for attitude statements like our original (2) and (3) to differ in truth-value. The thought is that an agent who stands in the believing relation to

the ILF of ‘Hesperus rises in the evening’ need not *ipso facto* stand in the believing relation to the ILF of ‘Phosphorus rises in the evening.’ Yet one who believes that Hesperus rises believes the very same semantic content as one who believes that Phosphorus rises.

Notice that Larson and Ludlow walk the middle ground between Davidson and Frege. Their approach is ontologically conservative like Davidson’s. For the structural descriptions and semantic values out of which ILFs are composed must already be introduced in to our ontology to account for the semantic and syntactic character of sentences other than those about propositional attitudes. But their explanation of failures of substitutivity is analogous to Frege’s own. They too appeal to the fact that, at least in attitude contexts, more is relevant to substitutivity than mere co-reference. That more is, however, not anything semantic. So we are not required, merely because of the peculiarities propositional attitude statements, to posit for each referring expression a second semantic role over and above its role of standing for a referent.

5 Hidden indexicality

Consider the following three theses:

- (a) *The Direct Reference Thesis*. The sole semantic role of a proper name is to stand for its bearer.
- (b) *Semantic Innocence*. A name plays the same semantic role when embedded within a that-clause as it plays when it occurs outside a that-clause.
- (c) *The Fregean Intuition*. Propositional attitude statements which differ only by co-referring proper names may differ in truth-value.

Each of these theses has considerable independent plausibility. But they are not obviously consistent. Frege was moved to abandon both thesis (a) and (b) because he thought they could not be reconciled with (c) (see [Proper names §5](#)). More recently, Nathan Salmon (1986) has abandoned (c) out of a commitment to (a) and (b). He has insisted that it is merely a pragmatically generated illusion that sentences like (2) and (3) have different truth-values. He claims that strictly, literally speaking, one who believes that Hesperus rises *ipso facto* believes that Phosphorus rises. It is just that it is conversationally misleading to talk that way. That is because belief is not a two place relation between a believer and a proposition, but a three place relation between a believer, a proposition and what we might call a ‘way of believing’ that proposition. Propositions are never believed *simpliciter*, but always in some way or other, under some guise or other. Moreover, Salmon insists, a rational cognizer can believe a given proposition under one guise while failing to believe it under a different guise. But the crucial further point is that nothing in an attitude ascription semantically specifies the guise under which the proposition named by the that-clause is supposed to be believed. The guise simply does not enter into the strict literal truth-conditions of an attitude statement. Nonetheless, Salmon suggests, we can pragmatically suggest something about the guise, by using one expression rather than another semantically equivalent one in making an attribution. If we use ‘Hesperus’ we suggest something different about the guise than we would if we were to use ‘Phosphorus’. That is why it is pragmatically misleading to use ‘Hesperus’ and ‘Phosphorus’ interchangeably in making attitude ascriptions.

By contrast, hidden indexicalists insist on all of our principles, and offer a strategy, with several variants, for reconciling them (see [Schiffer 1979](#); [Crimmins and Perry 1989](#); [Richard 1990](#); [Crimmins 1992](#)). Proponents of the hidden indexical approach generally agree with the view that belief is a three-place, rather than merely a two-place relation. They even agree that that-clauses which differ only by co-referring proper names must refer to the same (singular) proposition. But they insist that attitude ascriptions do not just pragmatically suggest information about the guise under which the specified proposition is believed. Information about the guise is part of what is strictly, literally said by an attitude ascription. That is why, against Salmon, the Fregean intuition is here to stay.

Yet, given the assumption of semantic innocence, how could that-clauses which differ only by corresponding co-referring proper names specify distinct ways of believing? The answer, according to the hidden indexicalist, is that there is more to an attitude statement than meets either the eye or the ear. Attitude reports contain hidden or unarticulated indexical constituents. These unarticulated constituents are not reflected in the surface grammatical structure of the sentence, but in a context of utterance they somehow refer to a way of believing. Consequently, attitude sentences, in context, manage to express propositions with more explicit constituents than they themselves have. Such sentences can be compared, the hidden indexicalist says, to sentences like the following in which one term of a relation is suppressed, but can be supplied, in context, by competent producers and consumers:

It is raining.	(Where?)
I am ready.	(For what?)
That is tall.	(For a what?)
Smith believes that <i>p</i> .	(How?)

When one utters, ‘It is raining’ on a certain occasion, one strictly, literally says that it is raining in a certain place - typically the place of utterance - even though no explicit constituent of one’s utterance refers to that place. Just so, the hidden indexicalist maintains, when one utters a belief report on an occasion, one strictly, literally says that a certain proposition is believed in a certain way. And one does so even though no explicit constituent of one’s utterances refers to the way of believing. Now the crucial further claim has to be that although ‘Hesperus’ and ‘Phosphorus’ have the same semantic value and function innocently, even in that-clauses, nonetheless, exchanging one for the other can affect the way an unarticulated constituent gets hooked up to a reference as a function of context. And that can explain why they are not freely substitutable one for another.

Clearly, the hidden indexicalist owes us a detailed account of these matters. We need to be told what a way of believing is. We need to know how specified ways of believing vary as a function of facts about context and facts about articulated constituents. Fortunately, the best advocates of this approach have gone a long way towards answering such questions.

See also: [Demonstratives and indexicals](#); [Indirect discourse](#); [Propositional attitudes](#)

KENNETH A. TAYLOR

References and further reading

- Carnap, R.** (1947) *Meaning and Necessity*, Chicago, IL: University of Chicago Press; enlarged 2nd edn, 1956. (Introduces the notion of an intensional isomorphism, the earliest version of the structured complex approach.)
- Cresswell, M.** (1984) *Structured Meanings*, Cambridge, MA: MIT Press. (Argues for structured intensions as objects of propositional attitudes.)
- Crimmins, M.** (1992) *Talk about Beliefs*, Cambridge, MA: MIT Press. (Defends a hidden indexical theory.)
- Crimmins, M. and Perry, J.** (1989) ‘The Prince and the Phone Booth: Reporting Puzzling Beliefs’, *Journal of Philosophy* 60: 435-45. (Defends a hidden indexical theory.)
- Davidson, D.** (1969) ‘On Saying That’, in D. Davidson and J. Hintikka (eds) *Words and Objections: Essays on the Work of W.V. Quine*, Dordrecht: Reidel, 158-74. (Articulates the paratactic analysis of indirect discourse.)
- Davidson, D.** (1967) ‘Truth and Meaning’, *Synthese* 17: 304-23. (Articulates Davidson’s broader semantic programme.)
- Devitt, M.** (1996) *Coming to our Senses*, Cambridge: Cambridge University Press. (Defends a neo-Fregean account of attitude statements without positing Fregean senses; rejects hidden indexical approaches and direct reference.)
- Fodor, J.** (1989) ‘Substitution Arguments and the Individuation of Beliefs’, in G. Boolos (ed.) *Method, Reasons, and Language*, Cambridge: Cambridge University Press. (Construes modes of presentations as expressions in the language of thought, while holding a referentialist semantics.)
- Frege, G.** (1892) ‘Über Sinn und Bedeutung’ (On Sense and Reference), *Zeitschrift für Philosophie und philosophische Kritik* 100: 25-50. (A seminal work of Frege’s in which he argues for the distinction between the sense and reference of expressions and presents a number of important theses regarding the two notions. Translated as ‘On Sense and Meaning’ in P.T. Geach and M. Black (eds) *Translations from the Philosophical Writings of Gottlob Frege*, Oxford: Blackwell, 3rd edn, 1980.)
- Higginbotham, J.** (1991) ‘Belief and Logical Form’, *Mind and Language* 6: 344-69. (Defends a version of the structured complex approach, drawing on the resources of contemporary syntactic theory. Relates this approach to Davidsonian extensional semantics.)
- Larson, R. and Ludlow, P.** (1993) ‘Interpreted Logical Forms’, *Synthese* 95: 305-57. (Articulates and defends the ILF approach.)
- Quine, W.V.** (1956) ‘Quantifiers and the Propositional Attitudes’, repr. in *The Ways of Paradox and Other Essays*, Cambridge, MA: Harvard University Press, revised and enlarged edn, 1976. (This widely reprinted article contains Quine’s argument against the permissibility of quantifying into opaque contexts.)
- Quine, W.V.** (1960) *Word and Object*, Cambridge, MA: Harvard University Press. (Forcefully articulates Quine’s disdain for intensions as ‘creatures of darkness’.)

- Recanati, F.** (1993) *Direct Reference*, Oxford: Blackwell. (A full-scale defence of the direct reference thesis which articulates yet another way to reconcile direct reference, innocence and the Fregean intuition.)
- Richard, M.** (1990) *Propositional Attitudes: an Essay on Thoughts and How We Ascribe Them*, Cambridge: Cambridge University Press. (Argues that propositional attitude statements express relations to Russellian annotated matrices - a version of the structured complex approach; also defends the view that belief reports are indexical.)
- Salmon, N.** (1986) *Frege's Puzzle*, Cambridge, MA: MIT Press. (Denies the Fregean intuition and defends innocence and direct reference.)
- Schiffer, S.** (1979) 'Naming and Knowing', in P.A. French, T.E. Uehling, Jr and H.K. Wettstein (eds) *Contemporary Perspectives in the Philosophy of Language*, Minneapolis, MN: University of Minnesota Press. (First appearance of the hidden indexical approach.)
- Segal, G.** (1989) 'A Preference for Sense and Reference', *Journal of Philosophy* 86: 73-89. (Another defence of a structured complex approach.)
- Soames, S.** (1987) 'Direct Reference, Propositional Attitudes, and Semantic Content', *Philosophical Topics* 15: 47-87. (Denies the Fregean intuition, while defending innocence and direct reference.)

Propositions, sentences and statements

A sentence is a string of words formed according to the syntactic rules of a language. But a sentence has semantic as well as syntactic properties: the words and the whole sentence have meaning. Philosophers have tended to focus on the semantic properties of indicative sentences, in particular on their being true or false. They have called the meanings of such sentences ‘propositions’, and have tied the notion of proposition to the truth-conditions of the associated sentence.

The term ‘proposition’ is sometimes assimilated to the sentence itself; sometimes to the linguistic meaning of a sentence; sometimes to ‘what is said’; sometimes to the contents of beliefs or other ‘propositional’ attitudes. But however propositions are defined, they must have two features: the capacity to be true or false; and compositional structure (being composed of elements which determine their semantic properties).

One reason for distinguishing a sentence from ‘what the sentence says’ is that a sentence may be meaningless, and hence say nothing, yet still be a sentence. But perhaps the main reason is that two people, *A* and *B*, may utter the same sentence, for example, ‘I am hot’, and say the same thing in one sense but not in another. The sense in which they say the same thing is that they use the same words with the same linguistic ‘meaning’. The sense in which they say something different is that they put the same words to different ‘uses’: *A* uses ‘I’ to refer to *A*, while *B* uses it to refer to *B*. Hence what *A* says may be true although what *B* says is false. If what is said can be true in one case and false in another, *A* and *B* have not made the same ‘statement’. On the other hand, if *B* utters instead ‘You are hot’ when *A* utters ‘I am hot’, they put different sentences to the same use and make the same statement. On this view (see [Strawson 1952](#)), we must distinguish the ‘sentence’, the ‘use’ of the sentence, and the ‘statement’ made by using a sentence in a context of utterance. According to [Strawson](#), it is not sentences but statements that are true or false.

Logicians usually abstract from the context of utterance of sentences in actual communication and talk of the propositions expressed as abstract entities. The main modern proponent of this conception is Frege. For Frege ([1918](#)), a proposition is a ‘thought’, which is both the cognitive meaning expressed by a sentence and the content of a propositional attitude such as belief or desire; thoughts are the ‘senses’ of sentences. Thoughts are distinguished according to the following principle: if it is possible rationally to believe that *p* and not to believe that *q* then the thought that *p* and the thought that *q* are distinct (see [Frege, G. §§3-4](#); [Sense and reference §2](#)).

Under a different, but related, conception, proposed by Carnap ([1947](#)), the proposition expressed by a sentence *S* is the set of possible worlds in which *S* is true (see [Semantics, possible worlds §9-10](#)). This view violates the compositional structure requirement, since such propositions are not grasped by grasping their components (see [Compositionality](#)). It also seems unable to differentiate between some distinct propositions. For example, there would only be one necessarily false proposition, since there is only one empty set of worlds, but intuitively there are many different necessarily false propositions.

An alternative view of propositions as entities is the (neo-)Russellian view that they are collections of actual entities making up ‘facts’ or ‘states of affairs’ (see [Facts §1](#)). Thus the proposition expressed by ‘Socrates is mortal’ is the ordered pair $\langle \text{Socrates, being mortal} \rangle x$, composed of the individual Socrates and the property of being mortal. Three questions for this view are: (1) Are the contents of false sentences negative facts, and are there such facts? (2) What is the criterion of identity for facts? (3) How can propositions as facts be the contents of propositional attitudes? The last of these can be seen as a version of the problem which led Frege to postulate senses: for the fact that $a = b$ must be the same fact as $a = a$, although someone who believes that $a = a$ does not thereby believe that $a = b$, and hence the propositions are not the same (see [Sense and reference §1](#)). Fregeans conclude that this shows that we cannot dispense with the notion of sense. If Russellians deny this, they must complicate their account of facts or their account of propositional attitudes (see [Propositional attitude statements](#)).

Whether one defines the intuitive notion of ‘what is said’ as a context-independent entity or by recourse to the notion of a statement or proposition expressed by an utterance in a particular context, one must say what it means for two sentences to express ‘the same proposition’ or make ‘the same statement’. Both phrases rely on the notions of meaning and synonymy, criticized by Quine ([1968](#); see [Quine, W.V. §8](#)). But if we dispense altogether with

propositions, statements or any other notion of the content of what is said, and choose instead to take sentences as truth bearers, we face two problems. The first is that sentences are unlikely candidates for the role of contents of propositional attitudes: if ‘I believe I am stupid’ introduces a relation between me and the English sentence ‘I am stupid’, it should be translated into German (say) as ‘Ich glaube "I am stupid"', although the correct translation is ‘Ich glaube daß ich dumm bin’, which says nothing about any English sentence (Church 1950). The second problem is that, in spite of our ontological scruples about admitting propositions as entities, we still need such a notion to express the content of what is said or believed, and to account for the intentionality of thought in general. In that respect, however vague or ill-defined the notion of a proposition, it cannot be dispensed with.

See also: [Intensional entities](#); [Logical and mathematical terms, glossary of](#)

PASCAL ENGEL

References and further reading

- Blackburn, S.** (1975) ‘The Identity of Propositions’, in S. Blackburn (ed.) *Meaning, Reference and Necessity*, Cambridge: Cambridge University Press. (A good statement of a deflationary alternative to the ontological conception of the problem.)
- Carnap, R.** (1947) *Meaning and Necessity*, Chicago, IL: University of Chicago Press, 2nd edn, 1956. (The first modern formulation of the possible worlds account.)
- Cartwright, R.** (1962) ‘Propositions’, in R.J. Butler (ed.) *Analytical Philosophy, First Series*, Oxford: Blackwell, 81-103. (A classical discussion of the main issues.)
- Church, A.** (1950) ‘On Carnap’s Analysis of Statements of Assertion and Belief’, *Analysis* 10: 97-9. (States his translation argument against sentences as objects of belief.)
- Frege, G.** (1918) ‘Der Gedanke: eine logische Untersuchung’, *Beiträge zur Philosophie des deutschen Idealismus* 1: 58-77; trans. P.T. Geach and R. Stoothoff, ‘Thoughts’, in *Collected Papers*, ed. B.F. McGuinness, Oxford: Blackwell, 1984, 351-72. (The classic statement of propositions as ‘thoughts’ or abstract entities.)
- Geach, P.T.** (1965) ‘Assertion’, *Philosophical Review* 74(4); repr. in *Logic Matters*, Oxford: Blackwell, 1968. (A criticism of the notion of statement.)
- Nuchelmans, G.** (1973) *Theories of the Proposition*, Amsterdam: North Holland. (The best historical account, although somewhat doxographic.)
- Quine, W.V.** (1968) ‘Propositional Objects’, *Critica* 5: 3-22; repr. in *Ontological Relativity and Other Essays*, New York: Columbia University Press, 1969. (Includes a summary of Quine’s misgivings about propositions as entities.)
- Russell, B.A.W.** (1918) ‘The Philosophy of Logical Atomism’, in *Logic and Knowledge: Essays 1901-1950*, ed. R.C. Marsh, London: Routledge, 1992. (Treats propositions as states of affairs in the world.)
- Salmon, N. and Soames, S.** (eds) (1988) *Propositions and Attitudes*, Oxford: Oxford University Press. (A good anthology on the semantics of propositional attitudes, focused on the Russellian view.)
- Stalnaker, R.** (1976) ‘Propositions’, in A.F. MacKay and D.D. Merrill (eds) *Issues in the Philosophy of Language*, New Haven, CT: Yale University Press, 79-92. (A good account of the possible worlds view.)
- Strawson, P.F.** (1952) ‘On Referring’, *Mind* 59: 320-44; repr. in *Logico-Linguistic Papers*, London: Methuen, 1971. (Puts the classical arguments in favour of the notion of statement.)

Provability logic

Central to Gödel's second incompleteness theorem is his discovery that, in a sense, a formal system can talk about itself. Provability logic is a branch of modal logic specifically directed at exploring this phenomenon. Consider a sufficiently rich formal theory T . By Gödel's methods we can construct a predicate in the language of T representing the predicate 'is formally provable in T '. It turns out that T is able to prove statements of the form

(1) If A is provable in T , then it is provable in T that A is provable in T .

In modal logic, predicates such as 'it is unavoidable that' or 'I know that' are considered as modal operators, that is, as non-truth-functional propositional connectives. In provability logic, 'is provable in T ' is similarly treated. We write $\Box A$ for 'A is provable in T '. This enables us to rephrase (1) as follows:

(1') $\Box A \rightarrow \Box \Box A$.

This is a well-known modal principle amenable to study by the methods of modal logic.

Provability logic produces manageable systems of modal logic precisely describing all modal principles for $\Box A$ that T itself can prove. The language of the modal system will be different from the language of the system T under study. Thus the provability logic of T (that is, the insights T has about its own provability predicate as far as visible in the modal language) is decidable and can be studied by finitistic methods. T , in contrast, is highly undecidable. The advantages of provability logic are: (1) it yields a very perspicuous representation of certain arguments in a formal theory T about provability in T ; (2) it gives us a great deal of control of the principles for provability in so far as these can be formulated in the modal language at all; (3) it gives us a direct way to compare notions such as knowledge with the notion of formal provability; and (4) it is a fully worked-out syntactic approach to necessity in the sense of Quine.

1 The prehistory of the subject

The history of provability logic proper starts with Kurt Gödel (1933), who, at the end of a discussion of the interpretation of intuitionistic propositional logic, briefly (in six lines!) discusses formal provability as a modal operator.

The next step was taken by Leon Henkin who, in 1952, asked a seemingly frivolous question:

If Σ is any standard formal system adequate for recursive number theory, a formula (having a certain integer q as its Gödel number) can be constructed which expresses the proposition that the formula with Gödel number q is provable in Σ . Is this formula provable or independent in Σ ?

(Henkin 1952: 180)

In short, is a sentence that 'says' 'I am provable in Σ ' provable in Σ ? Note that $\underline{Q} = \underline{Q}$ is a fixed point of $\text{Prov}_T(x)$: $\underline{Q} = \underline{Q}$ is provable in T and so is $\text{Prov}_T(\#(\underline{Q} = \underline{Q}))$; hence, by classical logic,

$(\underline{Q} = \underline{Q} \leftrightarrow \text{Prov}_T(\#(\underline{Q} = \underline{Q})))$ is provable in T . (See §2 below for an explanation of the notation.) $\underline{Q} = \underline{Q}$ is,

however, not a *literal* Henkin sentence: it is not built explicitly to say of itself 'I am provable' in the way that Gödel's sentence is built to say 'I am not provable'. In 1953, G. Kreisel published a noteworthy paper answering Henkin's question with a resounding 'It depends...'. Kreisel constructed two extensionally correct provability predicates where for the one the literal Henkin sentence ('I am provable in Σ ') is provably true, but for the other it is provably false.

It might have ended there. Kreisel's result might have discouraged everyone from thinking further about Henkin's 'frivolous' question. Happily for the subject, a second paper, by M.H. Löb, appeared in 1955. Löb shows that under certain natural assumptions (the ones codified by (L1)-(L3) below) on the provability predicate, the answer to Henkin's question is that the literal Henkin sentence is true. In fact, under these conditions every solution to the 'Henkin equation' is provably true. With the formulation of the Löb conditions the modal study of provability could begin. For a good part of the subsequent story the reader is referred to Boolos and Sambin (1991).

2 What is provability logic?

Provability logic is best described as the sustained investigation of formal provability and related notions as modal operators. Consider the language L of modal propositional logic (see **Modal logic**). Let A be a language of predicate logic and let T be a theory in A . T should satisfy a number of conditions, which we will specify later. For the moment let us simply pretend that A is the language of arithmetic and that T is Peano arithmetic (PA). An ‘interpretation’ $(.)\sigma$ of L in A with respect to T is given by a mapping σ of the propositional variables to sentences of A . This mapping is extended to a full interpretation via the following clauses:

$$\begin{aligned}(p)\sigma &= \sigma(p), \\ (\perp)\sigma &= \perp, (\top)\sigma = \top, \\ (A \wedge B)\sigma &= (A)\sigma \wedge (B)\sigma, \\ (A \vee B)\sigma &= (A)\sigma \vee (B)\sigma, \\ (A \rightarrow B)\sigma &= (A)\sigma \rightarrow (B)\sigma, \\ (\Box A)\sigma &= \text{Prov}_T(\#((A)\sigma)).\end{aligned}$$

Only the last clause deserves comment. Prov_T is the arithmetized provability predicate for T . $(A)\sigma$ is a sentence of A . $\#((A)\sigma)$ is the Gödel number of $(A)\sigma$. Finally, $\#((A)\sigma)$ is the numeral of $\#((A)\sigma)$. Thus $\#((A)\sigma)$ is a closed term of A denoting $\#((A)\sigma)$ under the standard interpretation. For example, $(\neg\Box\perp)\sigma$ translates to $\neg\text{Prov}_T(\#(\perp))$, the sentence expressing the consistency of T . $\#((A)\sigma)$ is, in a sense, the obvious way available to T to talk about $(A)\sigma$.

The meaning of ‘ $(.)\sigma$ ’ depends on the given T . Strictly speaking we should write ‘ $(.)\sigma, T$ ’, or something similar, to make the dependence on T visible. However, to avoid too heavy a notational burden, we ‘hide’ T .

In the definition of $(A)\sigma$ a subformula occurrence of B in A can ‘go to’ two different things. If the occurrence is not in the scope of a ‘ $\Box$ ’, then B translates to a formula of A . If the occurrence is in the scope of a ‘ $\Box$ ’, then the translation of B contributes to the value of a term in $(A)\sigma$. This difference in treatment leads to the existence of modalized fixed points. Let us remind ourselves of the Gödel fixed point lemma (see **Gödel’s theorems**). For any formula $C(x)$ of A with at most x free, there is a sentence D of A , such that $T \vdash D \leftrightarrow C(\#(D))$. Observe that D ‘occurs’ in two different ways in the fixed point statement - as a formula and as a term. The fixed point lemma has the ‘fixed point theorem’ (1) below as a correlate in the modal language.

To formulate (1) we first give a definition. We say that p occurs ‘only modalized’ in A if all occurrences of p in A are in the scope of a box. For example, p occurs only modalized in $\Box(p \rightarrow \Box\Box p)$; p does not occur only modalized in $p \wedge \Box p$. Note that if p does not occur in A at all, then it occurs only modalized in A .

(1) Suppose that p occurs only modalized in A . Consider any σ from the propositional atoms to the sentences of A . We can find a D in A such that if τ is the function such that $\tau(q) = \sigma(q)$ for $q \neq p$ and $\tau(p) = D$, then we have $T \vdash (p \leftrightarrow A)\tau$.

For example, we can find a τ such that $T \vdash (p \leftrightarrow \neg\Box p)\tau$. Note that this is just a reformulation of the most celebrated application of the Gödel fixed point lemma: the proof of the existence of a Gödel sentence G such that $T \vdash G \leftrightarrow \neg\text{Prov}_T(\#(G))$.

Note that if we were to drop the ‘guardedness condition’, that p must be only modalized, then we would immediately get the liar paradox by taking $A = \neg p$. We will see below that (1) has an unexpected improvement in (4).

Let Γ be a finite set of L -formulas. We write $(\Gamma)\sigma$ for $\{(A)\sigma \mid A \in \Gamma\}$. We define $\Gamma \models_T A$ (pronounced ‘ A is a T -consequence of Γ ’) to mean that for all σ ,

$$T + (\Gamma)\sigma \vdash (A)\sigma.$$

‘ $\models_T$ ’ is a relation of arithmetical validity (with respect to T) for the modal language. Interpretations take the place of semantics in this definition. The dependence on T in the definition is two-fold: first, we are considering

principles derivable in T ; second, the interpretation of ‘ $\Box$ ’ under the $(\cdot)\sigma$ is the arithmetization of the predicate ‘is provable in T ’.

We have, for example, $\Box p \vdash_T \Box\Box p$. Translating this back we find that it means that, for all sentences A of the language of T ,

$$T + \text{Prov}_T(\underline{\#(A)}) \vdash \text{Prov}_T(\underline{\#(\text{Prov}_T(\underline{\#(A)})))}.$$

You could say that $\Box p \vdash_T \Box\Box p$ expresses the validity of a schematic step with respect to T . The unusual thing is that the schema is formulated in a different kind of language: the language of modal propositional logic, which contains a logical operator not occurring in the language of T .

The relation ‘ $\vdash_T$ ’ can be axiomatized by a theory called Löb’s logic (L or PrL for provability logic, or GL for Gödel-Löb logic) with the following rules of proof:

- (L1) $\vdash A \Rightarrow \vdash \Box A$
- (L2) $\vdash \Box(A \rightarrow B) \rightarrow (\Box A \rightarrow \Box B)$
- (L3) $\vdash \Box A \rightarrow \Box\Box A$
- (L4) $\vdash \Box(\Box A \rightarrow A) \rightarrow \Box A$.

(L4) is also known as Löb’s principle. In view of the fact that in classical logic an implication follows from its consequent, Löb’s principle tells us that an instance of the reflection principle $\vdash \Box A \rightarrow A$ is *only* provable if its consequent is already provable. In other words, T ‘knows’ that it can prove only trivial instances of reflection.

The soundness of L was, in effect, proved by M.H. Löb (1955). Completeness was proved by R. Solovay (1976). ‘Solovay’s completeness theorem’ is as follows:

(2) Under appropriate conditions on the theory T (see below), $L + \Gamma \vdash A \Leftrightarrow \Gamma \vdash_T A$.

L has a Kripke-type completeness theorem with respect to transitive models with an upwardly well-founded accessibility relation. (The accessibility relation is upwardly well-founded if there are no infinite ascending chains.) Since the logic has the finite model property, we have Kripke completeness also with respect to finite, transitive, irreflexive models. The logic is not compact for the Kripke semantics: one can produce an infinite set X of modal formulas that cannot be valid on any transitive, upwardly well-founded model. On the other hand, for every finite subset X_0 of X one can find such a model on which X_0 is valid.

There are two major results about what L can prove. The first (due to Bernardi, de Jongh and Sambin) is the fact that fixed points of modalized formulas are unique. The theorem says that if p and q are truly and provably fixed points of A , then they are truly and provably equivalent. We write $A(B/p)$ for the result of substituting B for p in A . Define $\Box^+ A := A \wedge \Box A$.

(3) Suppose that p occurs only modalized in A and that q does not occur in A . Then

$$\Box^+(p \leftrightarrow A), \Box^+(q \leftrightarrow A(q/p)) \vdash \Box^+(p \leftrightarrow q).$$

The second major result (due to de Jongh and Sambin) is the existence of explicit fixed points. The theorem says that the language L is rich enough to contain explicit definitions of the fixed points of formulas (with respect to propositional variables which occur only modalized in them).

(4) Suppose that p occurs only modalized in A . Then there is a B in L containing only variables occurring in A , but not containing p , such that $\vdash B \leftrightarrow A(B/p)$.

The proof of (4) yields an algorithm for computing B from A and p .

To illustrate (1), (3) and (4), consider $\neg\Box p$. (1) tells us that in the arithmetical interpretation this formula has a fixed point: precisely the celebrated Gödel sentence G . (3) tells us that, modulo provable equivalence in T , this sentence is unique. (4) tells us that we can find an explicit equivalent of G in the modal language. If we apply the algorithm associated to (4), we find that the explicit fixed point is $\neg\Box\perp$, the sentence expressing the consistency of T . Thus (4) links the unprovable sentence of Gödel’s first incompleteness theorem to the unprovable sentence of

Gödel's second incompleteness theorem (see [Gödel's theorems](#)).

The difference between (1) and (4) is as follows. (1) tells us that for all interpretations of the propositional variables (except p) of A (with p modalized), we can assign a sentence to p yielding the fixed point property. *Prima facie* this sentence need not be expressible in the modal language. (4) tells us that the modal language is rich enough to define the fixed point explicitly.

We close this section with some remarks about the precise choice of the theory T and A , the language of T . To make the results above work, T must be rich enough to do arithmetization in. This means that a sufficiently rich theory of arithmetic must be interpretable in T . The most natural choice for this theory is Buss' S_2^1 , a theory far weaker than such well-known theories as primitive recursive arithmetic, $I\Sigma_1^0$ or PA. The theory S_2^1 is not contained in $I\Delta_0$. S_2^1 has induction restricted to the Σ_1^b -formulas, a class connected to the logical study of feasible algorithms (see [Complexity, computational](#)). $I\Delta_0$ has induction restricted to the Δ_0 -formulas. In S_2^1 some functions are provably total which are not provably total in $I\Delta_0$. It turns out that the presence of S_2^1 is sufficient to verify the validity of L. However, to prove Solovay's completeness theorem (2) we seem to need more. Inspection of Solovay's proof shows that the theory we (*prima facie*) need is $I\Delta_0 + \text{Exp}$. Here Exp is an axiom expressing that the exponentiation function is total. It is consistent with everything we know that S_2^1 has a provability logic that is stronger than L! We can now give the first condition that the theory T should satisfy to be sound and complete for Löb's logic.

Condition 1: $I\Delta_0 + \text{Exp}$ is interpretable in T .

We do not demand that $I\Delta_0 + \text{Exp}$ be literally a sub-theory of T , since we want the results of provability logic to be applicable to such theories as ZF (see [Set theory](#)) in which arithmetic is present only via a translation of the arithmetical language into the language of set theory.

A second condition on the theory is that it be sufficiently sound. For example, by the second incompleteness theorem $U := \text{PA} + \text{Prov}_{\text{PA}}(\#(\perp))$ is consistent. As is easily seen, $U \vdash \text{Prov}_U(\#(\perp))$. So $\models_U \Box \perp$ and, hence, Solovay's completeness theorem cannot hold for U . We exclude such theories as U by demanding that T is Σ_1 -sound, that is, the Σ_1 -sentences T proves are all true. (A somewhat sharper condition is possible, but space does not permit its introduction here.)

Condition 2: T is Σ_1 -sound.

3 What does provability logic achieve?

We end with some comments on the usefulness of provability logic. First the mathematical side. Clearly the modal presentation of arithmetical arguments makes them very perspicuous. So provability logic achieves Kreisel's excellent aim of 'saying simple things simply'. Also, by Solovay's theorem, provability logic completely describes the arithmetically valid modal provability principles. Moreover, it enables us to study these principles using the tools of modal logic (for example, Kripke semantics). Thus, by switching to the modal language, we gain control.

Now for philosophical relevance. Provability logic offers a fully worked-out syntactic treatment of modality (in the sense of [Quine 1953](#)). Since its principles are quite different from what one would ordinarily expect of notions of knowledge, necessity and so on, one way of viewing what is achieved is to conclude that syntactic treatments of such notions are not possible (for a brief discussion, see [Thomason 1980](#)). Moreover, comparing intuitively plausible modal principles for notions such as knowledge, human provability and the like with the principles satisfied by formal provability gives some sense of how these notions differ from formal provability. The reflection principle $\vdash \Box A \rightarrow A$ is built into or internal to the ordinary notions of knowledge and (human) provability. In contrast, formal provability satisfies Löb's principle ($\vdash \Box(\Box A \rightarrow A) \rightarrow \Box A$). We can 'see' that the difference between human and formal provability is not simply one of *strength* (but, I would suggest, one of *kind*). (The popularity of the view that the difference is one of strength can be seen, for example, in the discussion concerning minds and machines, where it is sometimes claimed on the ground of Gödel's incompleteness theorems that humans can beat machines in a theorem-producing game.)

See also: [Logical and mathematical terms, glossary of](#)

ALBERT VISSER

References and further reading

- Boolos, G.** (1993) *The Logic of Provability*, New York and Cambridge: Cambridge University Press. (A very readable textbook; includes an up-to-date introduction to predicate provability logic.)
- Boolos, G. and Sambin, G.** (1991) 'Provability: The Emergence of a Mathematical Modality', *Studia Logica* 50 (1): 1-23. (A brief history of the subject for roughly the period 1933-80.)
- Gödel, K.** (1933) 'Eine Interpretation des intuitionistischen Aussagenkalküls', *Ergebnisse eines Mathematischen Kolloquiums* 4: 39-40; trans. 'An Interpretation of the Intuitionistic Propositional Calculus', in *Collected Works*, ed. S. Feferman, J.W. Dawson, W. Goldfarb, C.D. Parsons and R.M. Solovay, Oxford and New York: Oxford University Press, 1995, vol. 3, 296-302. (Includes a thorough discussion by A.S. Troelstra.)
- Henkin, L.** (1952) 'A Problem concerning Provability', *Journal of Symbolic Logic* 17: 160, problem 3. (The paper in which Henkin posed the crucial question 'Is a sentence that "says" "I am provable in Σ " provable in Σ ?'.)
- Kreisel, G.** (1953) 'On a Problem of Henkin's', *Indagationes Mathematicae* 15: 405-6. (Kreisel's answer to Henkin: 'It depends...'.)
- Löb, M.H.** (1955) 'Solution of a Problem of Leon Henkin', *Journal of Symbolic Logic* 20: 115-18. (As the title of this paper suggests, Löb answered Henkin's problem, and also proved the soundness of Löb's logic.)
- Quine, W.V.** (1953) 'Three Grades of Modal Involvement', in *Proceedings of the 11th International Congress of Philosophy* 14, Amsterdam: North Holland, 65-81. (Quine shows that provability logic offers a fully worked-out syntactic treatment of modality.)
- Smoryński, C.** (1985) *Self-Reference and Modal Logic*, New York: Springer. (A well-written and elegant textbook; includes good discussions of multi-modal provability logic and of non-extensional self-reference.)
- Solovay, R.M.** (1976) 'Provability Interpretations of Modal Logic', *Israel Journal of Mathematics* 25: 287-304. (A proof of the completeness of Löb's logic.)
- Thomason, R.H.** (1980) 'A Note on Syntactical Treatments of Modality', *Synthese* 44: 391-5. (Presents the view that syntactic treatments of knowledge, necessity and so on are not possible.)

Pythagoreanism

Pythagoreanism refers to a Greek religious-philosophical movement that originated with Pythagoras in the sixth century BC. Although Pythagoreanism in its historical development embraced a wide range of interests in politics, mysticism, music, mathematics and astronomy, the common denominator remained a general adherence among Pythagoreans to the name of the founder and his religious beliefs. Pythagoras taught the immortality and transmigration of the soul (reincarnation) and recommended a way of life that through ascetic practices, dietary rules and ethical conduct promised to purify the soul and bring it into harmony with the surrounding universe. Thereby the soul would become godlike since Pythagoras believed that the cosmos, in view of its orderly and harmonious workings and structure, was divine. Pythagoreanism thus has from its beginnings a cosmological context that saw further evolution along mathematical lines in the succeeding centuries. Pythagorean philosophers, drawing on musical theories that may go back to Pythagoras, expressed the harmony of the universe in terms of numerical relations and possibly even claimed that things are numbers. Notwithstanding a certain confusion in Pythagorean number philosophy between abstract and concrete, Pythagoreanism represents a valid attempt, outstanding in early Greek philosophy, to explain the world by formal, structural principles. Overall, the combination of religious, philosophical and mathematical speculations that characterizes Pythagoreanism exercised a significant influence on Greek thinkers, notably on Plato and his immediate successors as well as those Platonic philosophers known as Neo-Pythagoreans and Neoplatonists.

1 History

In the second half of the sixth century BC, Pythagoras founded a community in the southern Italian city of Croton whose members were united by the belief in the transmigration of the soul, an ascetic way of life that centred on the purification of the soul, and a political outlook that aimed at social reform along aristocratic lines. Pythagorean associations (*hetaireiai*), which also formed in other cities of Magna Graecia, acquired considerable political authority, but their dominance eventually met with opposition, both during the lifetime of Pythagoras and later again about 450 BC. In the wake of these anti-Pythagorean movements the followers of Pythagoras were scattered throughout the Greek world, so that by the time of Plato there is little evidence of formal Pythagorean societies. Individual Pythagoreans, however, continued to be recognized, some by their distinctive lifestyle in matters of food, dress and purificatory practices, others by the additional pursuit of various philosophical, mathematical and musical theories with which they credited Pythagoras. Later tradition refers to these two types of Pythagoreans as ‘hearers’ (*akousmatikoi*) and ‘learners’ (*mathēmatikoi*). The distinction supposedly goes back to the original society in which some members were only fit to accept the oral teachings of Pythagoras without arguments and proofs, while those with more leisure and perhaps philosophical ability were further instructed in the rational foundations of the master’s teachings. Whatever differences there may have been between groups of Pythagoreans, and even among individual *mathēmatikoi*, all professed allegiance to Pythagoras.

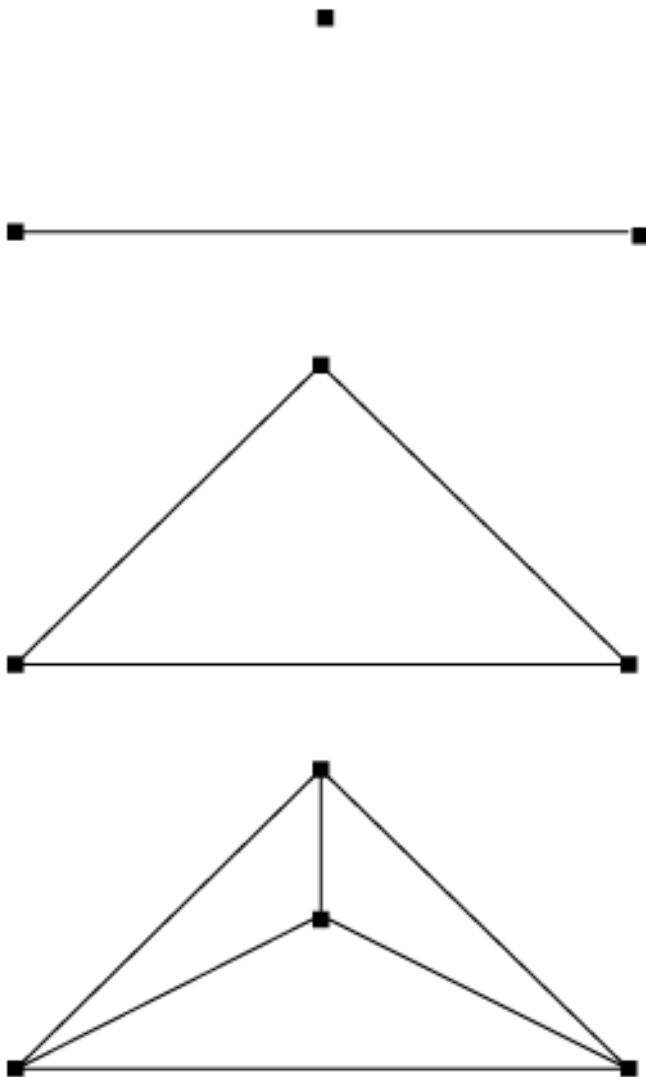
From the time of Pythagoras to Plato there were several famous Pythagoreans - Hippasus, **Philolaus** and **Archytas**. Aristotle, in his extant works, speaks only generally of the Pythagoreans (‘some Pythagoreans say...’); his special treatise on Pythagorean beliefs unfortunately no longer survives. In the third and second centuries BC we do not hear of philosophers who were known as or called themselves Pythagoreans, but interest in ‘Pythagoreanism’ continued, as is evidenced by the wealth of apocryphal writings in prose and verse on Pythagorean themes that mostly date from this period. The actual practice of Pythagoreanism experienced a revival in the Roman world from the first century BC to the first century AD; Latin writers such as Cicero, Ovid and Seneca testify to its popularity. In the first two centuries AD the theoretical side of Pythagoreanism marked certain philosophers to the extent that they may be called Neo-Pythagoreans (see **Neo-Pythagoreanism**) and these in turn influenced later Platonic philosophers (see **Neoplatonism**).

2 Music, mathematics, and cosmology

Plato says of the true philosopher, whose mind is on the higher realities (that is, the Platonic Forms):

he looks unto the fixed and eternally immutable realm where... all is orderly and according to reason, and he imitates this realm and, as much as possible, assimilates himself to it... and by association with the divine order becomes himself orderly and divine as far as a human being can...

Plato's philosophic ideal, as well as one of the methods he prescribed to achieve it - namely mathematics - owes much to Pythagoreanism. Pythagoras had taught that life should be in harmony with the divine cosmos (see [Pythagoras §2](#)). In Pythagoreanism harmony (*harmonia*) became a central tenet and was explained through numerical relations, possibly in connection with musical theory. For example, the Pythagoreans thought that the motions of the orbiting planets produced a sound which, given the belief that the intervals between the heavenly bodies corresponded to musical ratios, was harmonious. So Aristotle explains the 'music of the spheres' - one of several explanations offered for this famous Pythagorean image. Of seminal importance for illustrating the coherence of music and number was the discovery of musical ratios - that music should result when the first four integers of the numerical system, used as components in the harmonic ratios of the octave (2 : 1), the fifth (3 : 2) and the fourth (4 : 3), were imposed upon the continuum of sound. Whether or not Pythagoras, as tradition holds, was the 'discoverer' of the musical concords, the Pythagoreans fixed upon the first four numbers as the building blocks of nature. These four sufficed to give extension and shape to bodies in the sequence of [point-line-surface-solid](#):

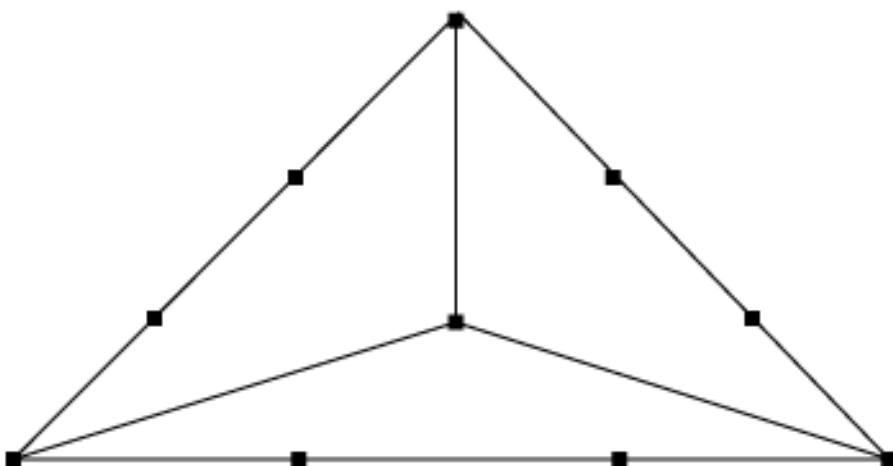
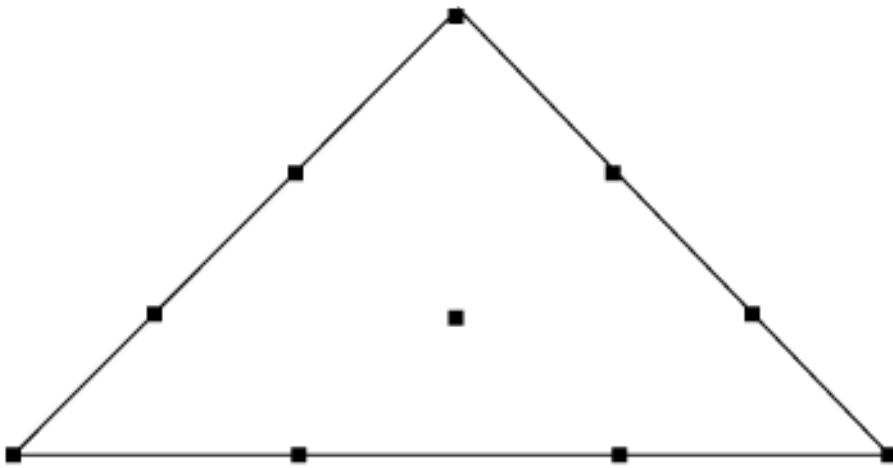


Moreover, the first four integers add up to ten, which the Pythagoreans considered a perfect number and represented in a [figure](#) called the *tetraktys*.

Pythagoreanism



From whatever angle one approached the *tetraktys* in counting, the addition always resulted in ten, the basis of the decimal system. The *tetraktys* also produced an equilateral triangle, the simplest plane figure, and a pyramid, the simplest three-dimensional shape.



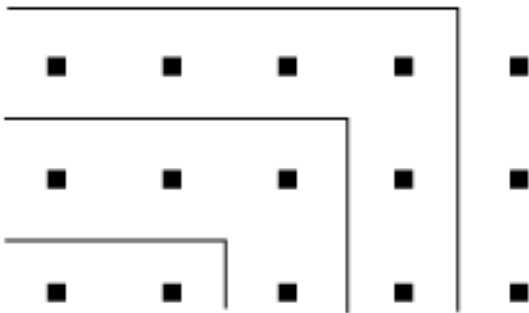
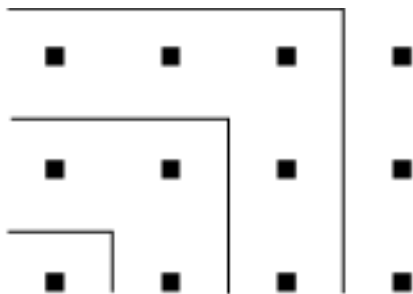
The Pythagoreans viewed the *tetraktys* as a sacred symbol and used it in the following oath: ‘By him [Pythagoras] who handed down to us the *tetraktys*, source and root of everlasting nature’. In short, the Pythagoreans supposed the nature of things could be understood numerically; indeed, some of them apparently went so far as to say that things *are* numbers: ‘...they assumed the elements of numbers to be the elements of all things that exist, and the whole universe to be *harmonia* and number’ (Aristotle, *Metaphysics* 986a2).

Since numbers functioned as the constituent elements of the cosmos, it is important to see how the Pythagoreans understood the nature and generation of number itself. Aristotle again offers the best starting-point:

The elements of number are the even and the odd, the latter limited, the former unlimited. The One is composed of both of these (for it is both even and odd) and number comes from the one; and numbers... are the whole universe.

(*Metaphysics* 986a17)

How the One can be both odd and even is best explained in the sense that the One, as the first number, is the principle of both odd and even numbers (zero was unknown in Greek mathematics). Thus ‘number comes from the One’, and the generation of number was simultaneously a cosmogonical process since ‘numbers... are the whole universe’. Although in its origin the Pythagoreans viewed the One as both odd and even, limited and unlimited, in its practical application, that is, in its interaction with other numbers, they treated it as odd and limiting. This can be seen in the schema by which the Pythagoreans illustrated the correspondence of odd/even to limited/unlimited. **Gnomons** (carpenter squares) were placed around an arrangement of points (or pebbles) as follows:



When a gnomon is placed around one point, and the process is continued in sequence, the resulting figure is always of ‘limited’ shape, that is, always a square, whereas when it is placed around two points, the result is a series of oblongs whose sides stand in an ‘unlimited’, that is, an infinite variation to each other. In this scheme the One, as a single unit or point, is equated with the odd. The One is also ranked with limited and odd in the Pythagorean table of ten opposites:

limited	unlimited
odd	even
one	many
right	left
male	female
rest	motion
straight	crooked
light	darkness
good	bad
square	oblong

The position of limited/unlimited at the beginning is not accidental, for their opposition, embodied in the original composition of the One, was considered primal and basic to the development of number and the universe, while the opposition between odd and even can be generally subsumed under this fundamental distinction. The connection of the One with the limited reappears in another Pythagorean text (Aristotle, [Metaphysics](#) 1091a17), where it is said that the limited (tantamount to the One) ‘breathes in’ and is thus penetrated by the unlimited. In this cosmological fragment the unlimited is equated with the void, to be thought of as infinite space and serving as a dividing principle. The One now becomes a Two (reminiscent of the mythological separation of heaven and earth), which in effect marks the beginning of number, plurality and the existence of discrete physical bodies. The cosmogonic process is sometimes put the other way round: the unlimited is limited (or penetrated) by the limit. This way of stating it brings to the fore the characteristic Greek feeling that what is unlimited, without bounds, is without order and somehow evil (hence ‘bad’ is listed under the unlimited in the above table) and therefore needs to be curtailed and bounded by limit and measure (as the limits of harmonic ratios must be imposed upon the limitless range of sound to produce music). The result, in cosmological terms, is then precisely a *kosmos*, an orderly and structured universe. In so far as the perfected and integral cosmos represents a *harmonia*, a unity stemming from a reconciliation of opposites, Pythagoreanism has a monistic aspect, yet in its source theory it remains essentially dualistic: number comes from the One, and the One is composed of limited and unlimited. From these first two principles everything derives, even though the One, when identified with limit, odd, male, and so on, was felt to be the ‘good’ element in this cosmic dualism. (There can be no question of positing the One as the sole ultimate principle without imposing Platonic and Neo-Pythagorean notions upon early Pythagoreanism.)

Limited and unlimited are the principles of number. Numbers do not merely express the substance, shape and quantitative differences of things but actually appear to constitute bodies of physical dimension. The identification of things with numbers confuses physical bodies with abstractions, or, in Aristotelian terms, the material causes of things with their formal causes. Here Pythagoreanism connects with much of Presocratic thought according to which the source and informing element of the universe was thought of as some kind of matter, be it water, air, fire, earth, or a combination of these (see [Archē](#)). At the same time the numerical theories of the Pythagoreans represent a true advance in the history of Greek philosophy: the attempt to explain the nature of things by numbers is a valid philosophical striving to understand the world by its formal or structural principles, even if the Pythagoreans then equated these with the things themselves. And while the Pythagoreans’ interest in numbers was often infused with a mystical element (their solemn veneration of the *tetraktys*) and a primitive number symbolism that blurred distinctions between abstract and concrete (a moral concept such as justice was considered the embodiment of four, a square number of ‘just’, that is, equal reciprocity), no less an authority than Aristotle acknowledged that the Pythagoreans were pioneers and made advances in the mathematical fields (arithmetic, astronomy, harmonics and some geometry). For the ‘Pythagorean theorem’, although long known to Babylonian mathematicians, the Pythagoreans are generally considered to have found the proofs and, although in the wake of this theorem they discovered that the ratios of geometrical figures to each other cannot simply be expressed by a series of rational integers - the discovery of irrational numbers and the principle of incommensurability upset the Pythagorean notion that the world is harmony and number - their very setbacks contributed to subsequent work in Greek mathematics. Plato valued mathematics as a useful discipline to train the philosopher in the perception of eternal and transcendent truth, since in his mind it dealt essentially with invisible and eternally valid realities. By divorcing number from physical substance Plato transforms Pythagorean mathematics, yet few doubt that his

interest in number, arithmetic and measure largely rests upon Pythagorean foundations (consider the mathematical model of the universe in his *Timaeus* (see [Plato §16](#)). Plato's immediate successors in the Academy were steeped in Pythagorean number theory (see [Speusippus §2](#); [Xenocrates §2](#)).

3 Soul and ethics

Perhaps in no other ancient Greek philosophy is the human being as intimately linked to the cosmos as in Pythagoreanism. Although the correspondence between microcosm and macrocosm, the individual and the universe, is found in much of Greek thought, it received a particularly sharp outline in the Pythagorean view that by assimilation to the divine cosmos the self would come to reflect the cosmic order and harmony. The true self of every person was the soul (see [Psychē](#)), the essential element in the partnership of body and soul. Pythagoras' teaching that the soul survived the dissolution of the body and reappeared in other bodies was steadfastly adhered to throughout the history of Pythagoreanism, even if it was sometimes eclipsed by other interests. The doctrine of the immortality and transmigration of the soul translated into a practical and clearly defined way of life that combined ritual purity with high ethical standards and whose precepts were embodied in sayings known as *akousmata* and *symbola* (see [Pythagoras §2](#)). The belief in the kinship of all life, as a corollary to the belief in transmigration, imposed a great moral responsibility towards parents, children, friends and fellow citizens, and generally entailed a respectful attitude towards all forms of life in which soul may be embodied; hence the Pythagoreans were, to varying degrees, vegetarians. Their dietary laws were also intended to free them from bodily pollution. The body was seen as a prison, even as a grave (Plato, *Gorgias* 493a) from which the soul was to rise, ever achieving superior reincarnations and culminating in a state of divinity. Thus the Pythagoreans postulated three kinds of rational creatures: 'gods, men, and such as Pythagoras'. Pythagoras was thought to have achieved semi-divine status, which in Greek could be expressed by 'daemon'. It is in a Pythagorean vein that [Empedocles \(§2\)](#) proclaims human beings to be fallen 'daemons'. (A complex 'daemonology' was to become an ingredient of Neo-Pythagoreanism.) The recognition of the religious and moral mandates of the Pythagorean life and of its eschatological meaning constituted wisdom (*sophia*) and the lover of such a life was a *philosophos*, a term that in this sense, as certain traditions report, was first coined in Pythagorean circles.

See also: [Mystical philosophy in Islam §2](#); [Orphism](#); [Presocratic philosophy](#)

HERMANN S. SCHIBLI

References and further reading

- Aristotle** (c. mid 4th century BC) *Metaphysics*, trans. in J. Barnes (ed.) *The Complete Works of Aristotle*, Princeton, NJ: Princeton University Press, 1984. (A primary source on early Pythagoreanism.)
- Burkert, W.** (1972) *Lore and Science in Ancient Pythagoreanism*, trans. E.L. Minar, Jr, Cambridge, MA: Harvard University Press. (A scholarly examination of the sources of Pythagoreanism and a critical analysis of Pythagorean teachings; a prerequisite for all serious work on the subject.)
- Diels, H and Kranz, W.** (eds) (1952) *Die Fragmente der Vorsokratiker (Fragments of the Presocratics)*, Berlin: Weidmann, 6th edn, vol. 1, 96-113 and 375-480. (The standard collection of the ancient sources containing the main sources on Pythagoreanism.)
- Guthrie, W.K.C.** (1962-78) *A History of Greek Philosophy*, Cambridge: Cambridge University Press, 6 vols. (A comprehensive and readable discussion of early Pythagoreanism can be found in volume 1, pages 146-340; emphasizes the harmony between its religious and philosophical-scientific elements.)
- Huffman, C.A.** (1993) *Philolaus of Croton: Pythagorean and Presocratic*, Cambridge: Cambridge University Press. (Examines the fragments of a notable Pythagorean; pages 57-64 argue that 'all things are number' was not an actual claim of the Pythagoreans but rather Aristotle's summary of their number theory.)
- Kahn, C.H.** (1974) 'Pythagorean Philosophy before Plato', in A.P.D. Mourelatos (ed.) *The Pre-Socratics*, Garden City, NY: Doubleday; repr. 1993. (A short but useful introduction.)
- Kingsley, P.** (1995) *Ancient Philosophy, Mystery and Magic: Empedocles and Pythagorean Tradition*, Oxford: Oxford University Press. (Examines Pythagoreanism against the background of Greek mystery and magic; also discusses the influence of Pythagoreans on the myths in Plato.)
- Navia, L.E.** (1990) *Pythagoras: An Annotated Bibliography*, New York and London: Garland. (Includes the literature, up to 1989, not only on Pythagoras but also on all aspects of Pythagoreanism.)
- Philip, J.A.** (1966) *Pythagoras and Early Pythagoreanism*, Toronto, Ont.: Toronto University Press. (An examination, based principally on Aristotle, of the major doctrines of Pythagoreanism; in regard to cosmology, holds that the Pythagorean opposites are not the first principles of a dualistic cosmos, but rather, given a

creation in time, the constituents of one *physis*, one world-nature.)

Plato (c.380-367 BC) *Republic*, trans. P. Shorey, *Plato: Republic*, Loeb Classical Library, Cambridge, MA: Harvard University Press and London: Heinemann, 1930, 2 vols.(Much Pythagorean influence in books VI-VII.)

Quantification and inference

Quantificational reasoning in natural languages contains occurrences of expressions which, though name-like from a syntactic perspective, intuitively seem to be importantly different from ordinary names. For example, in attempting to argue that every F is G , we might consider an 'arbitrary' F , give 'it' the 'name' ' n ', and go on to argue that n is G . That such an occurrence of ' n ' is somehow different from occurrences of ordinary names can be seen by noting, for example, that arguing from the claim that Bertrand Russell is F to the claim that Bertrand Russell is G in general does not suffice for drawing the conclusion that every F is G . Similarly, given that some F is G , in general we do not think it legitimate to assert that Bertrand Russell is F and G , and continue reasoning from that claim. It seems that such occurrences of ' n ' differ semantically from similar occurrences of ordinary names. Thus the question arises as to the semantics of occurrences of such expressions. An adequate semantic account must justify appropriate inferences to and from sentences containing these terms.

1 Condition of adequacy

Natural languages contain quantificational reasoning employing expressions which function syntactically, but not semantically, as proper names. For example, in attempting to argue that every F is G , we might consider an 'arbitrary' F , give 'it' the 'name' ' n ', and go on to argue that n is G . Or, having been given that some F is G , we give the 'name' ' n ' to an arbitrary F that is G , and go on to draw further conclusions concerning n . It will help to shed light on these terms in natural language arguments if we turn our attention to systems of natural deduction for first-order logic (see **Natural deduction, tableau and sequent systems**). For such systems contain rules which more or less mimic these natural language forms. Though formulations vary significantly, many systems have a 'universal generalization' (UG) rule which licenses the inference of ' $(\forall x)\phi(x)$ ' from ' $\phi(a)$ ' subject to certain restrictions, and an 'existential instantiation' (EI) rule which allows one to infer ' $\phi(a)$ ' from ' $(\exists x)\phi(x)$ ', again subject to restrictions. We shall call singular terms in systems of natural deduction which play the role played by ' a ' above 'instantial terms'. Instead of discussing terms such as ' n ' in our natural language examples, we can ask about the proper semantic treatment of instantial terms in systems of natural deduction. The regimentation and explicitness of structure of natural deduction will enable us to state conditions of adequacy precisely and to see whether those conditions are met by particular semantic theories.

As in natural language, instantial terms in natural deduction appear to behave differently from names or 'individual constants'. Both UG and EI fail to be truth-preserving if the instantial term is an individual constant. Of course, in many formulations of natural deduction, expressions from the syntactic category of individual constant (or individual variable) are pressed into service as instantial terms in applications of UG and EI. However, this probably results more from the desire not to have an additional category of singular terms than from any view about the appropriate semantics for instantial terms.

Before discussing conditions of adequacy for semantic theories of instantial terms, let us consider the view that there can be no semantic theory for instantial terms, because occurrences of formulas in applications of UG and EI which contain instantial terms should not be thought of as having any meaning. On this view, UG and EI have a certain utility in that they allow us to move from meaningful formulas to other meaningful formulas (and soundness and completeness results for systems employing these rules justify their use), but it is a mistake to think that formulas containing instantial terms in applications of UG and EI themselves have any meaning. Whatever attraction this view may have as an account of formulas containing instantial terms in formal systems, it seems quite implausible as an account of natural language sentences containing the analogues of these terms. For surely sentences containing such terms in natural language arguments are meaningful: as we read an argument containing these terms, we understand the claims made by the sentences containing them.

In an adequate semantics all allowable inferences must be truth-preserving. That is, the truth-conditions assigned to a formula must be such that, should they obtain, then so will the truth-conditions of any formula properly inferred from it. So in so far as we provide a semantics for instantial terms and the formulas containing them, it ought to account for the inferential relations such formulas manifest. Let us call this property of a semantics 'inferential adequacy'. It is important to appreciate the difference between inferential adequacy and classical completeness and soundness results. Quine's system of deduction ([1950] 1982) is illustrative. Quine defines a

‘finished derivation’ as a derivation whose last line neither contains an instantial term (a ‘flagged variable’, in his terminology) nor depends on a premise which contains an instantial term. He then proves what he calls soundness; that the premises imply the last line of a finished derivation. However, Quine provides no semantics for instantial terms that is inferentially adequate. As he himself remarks, UG and EI are not ‘implicative’ relative to the only semantics he gives (which essentially treats instantial terms as names).

An inferentially adequate semantics must assign distinct truth-conditions to occurrences of the same formula in different derivations. For suppose ‘ $(\forall x)\phi(x)$ ’ is correctly inferred from ‘ $\phi(a)$ ’ in some derivation. Inferential adequacy requires the truth-conditions of this occurrence of ‘ $\phi(a)$ ’ to be such that it entails ‘ $(\forall x)\phi(x)$ ’. On the other hand, an occurrence of ‘ $\phi(a)$ ’ in a derivation where it is correctly inferred from ‘ $(\exists x)\phi(x)$ ’ must be assigned truth-conditions such that it is entailed by the latter. Thus ‘ $\phi(a)$ ’ must have different truth-conditions in the two derivations.

There are different sorts of cases in which a given formula must be assigned different truth-conditions as it occurs in different derivations. Consider the following example.

(A)	(B)
(1) $(\forall x)(\exists y)Fxy$	(1) $(\exists y)(\forall x)Fxy$
(2) $(\exists y)Fay$	(2) $(\forall x)Fxb$
(3) Fab	(3) Fab
(4) $(\exists w)Faw$	(4) $(\forall z)Fzb$
(5) $(\forall z)(\exists w)Fzw$	(5) $(\exists w)(\forall z)Fzw$

In derivation (A), inferential adequacy requires that ‘ Fab ’ is logically equivalent to ‘ $(\forall x)(\exists y)Fxy$ ’, from which it is inferred. In (B) it must be logically equivalent to ‘ $(\exists y)(\forall x)Fxy$ ’. Note that this is unlike the previous case in that it is not simply a matter of (for example) ‘ $\phi(a)$ ’ being inferred from ‘ $(\exists x)\phi(x)$ ’ in one case and having ‘ $(\forall x)\phi(x)$ ’ inferred from it in another. In both (A) and (B) a formula containing ‘ a ’ has a universal generalization (on the place occupied by ‘ a ’) inferred from it.

2 Three semantic theories of instantial terms

Three semantic accounts of instantial terms have been shown to satisfy the requirement of inferential adequacy. Though other inferentially adequate accounts can be imagined, we shall consider only these three. The first holds that in addition to individual people, numbers and so on there are ‘arbitrary’ people, numbers and so on. (An arbitrary number has every property common to all numbers, and so on.) Instantial terms are names of these arbitrary or indefinite objects. Prior to Frege’s attack on such views (1904), they were widely held and are still sometimes suggested by informal comments accompanying the introduction of UG and EI in logic textbooks. Kit Fine (1985a) has made the view respectable by formulating a precise and powerful version that is inferentially adequate. On Fine’s formulation, which arbitrary object a given instantial term names is determined by features of the derivation in which it occurs. An arbitrary object is associated with a range of individual objects (an arbitrary number is associated with a range of individual numbers, and so on), and a predicate is true of the arbitrary object just in case it is true of each individual object in the range. It is this which allows ‘ $\phi(a)$ ’ to entail ‘ $(\forall x)\phi(x)$ ’ under the appropriate conditions.

A second account of the semantics of instantial terms holds that they are actually quantifiers. On this view, instantial terms are expressions of generality just like ordinary quantifiers except that the quantifications they express and the scope relations they have to each other are determined by linguistic context. So, for example, in (A) and (B) above, features of the derivations determine whether ‘ a ’ and ‘ b ’ express universal or existential quantifications, and whether ‘ a ’ or ‘ b ’ has wider scope (‘ a ’ has wide scope in (A) and narrow scope in (B) relative to ‘ b ’). On this view, suggested in Wilson (1984) and formulated as inferentially adequate in King (1991), we might say that instantial terms are ‘context-dependent quantifiers’, thus emphasizing both that they express quantifications and that the quantifications they express and the relative scopes of those quantifications vary with the linguistic context. On such a view, ‘ $\phi(a)$ ’ entails ‘ $(\forall x)\phi(x)$ ’ (under the appropriate conditions) because ‘ $\phi(a)$ ’ itself expresses a quantified claim.

The third view is that instantial terms are to be thought of as ε -terms. Hilbert introduced the ε -operator as a variable-binding operator used to form terms: if ' ϕ ' is a formula (some presentations require ' ϕ ' to contain free ' x '), then ' $\varepsilon x\phi$ ' is a singular term. Though Hilbert provided no explicit semantics for singular terms formed using the ε -operator, he remarked that ' $\varepsilon x\phi$ ' should be thought of as denoting an arbitrary element in the domain satisfying ' ϕ ' if anything does (if nothing satisfies ' ϕ ', the term can be thought of as denoting any element in the domain). Further, Hilbert provided the following axiom to govern ε -terms: $(\exists x)\phi(x) \rightarrow \phi(\varepsilon x\phi)$. This axiom justifies the following rules of inference for ε -terms: (1) infer ' $\phi(\varepsilon x\phi)$ ' from ' $(\exists x)\phi(x)$ '; and (2) infer ' $(\forall x)\phi(x)$ ' from ' $\phi(\varepsilon x\sim\phi)$ '. Thus Hazen (1987) considers constructing a system of natural deduction in which (1) and (2) replace EI and UG respectively, but in which we abbreviate the ε -terms used in applications of (1) and (2) by single letters. Derivations using such abbreviations would be indistinguishable from those in Quine's (1950) system. A semantics for ε -terms of the sort given in Leisenring (1969) would result in an inferentially adequate theory.

Which of the three views discussed above is ultimately preferable appears to depend on philosophical and methodological considerations. In fact, it is not clear how different the views really are. One might well complain that arbitrary objects are unnecessary intermediaries, and that the ranges of individuals with which they are associated do all the real semantic work. But when arbitrary objects are eliminated from the first view and instantial terms are directly associated with ranges of individuals, the first and second views begin to look quite similar. In addition, since, following Hilbert, we can use ε -terms to 'define' the quantifiers $(\forall x)\phi(x) \leftrightarrow \phi(\varepsilon x\sim\phi)$ and $(\exists x)\phi(x) \leftrightarrow \phi(\varepsilon x\phi)$, it is unclear whether the difference between the second view (instancial terms are quantifiers) and the third view (instancial terms are ε -terms) is substantive.

See also: [Logical and mathematical terms, glossary of](#)

JEFFREY C. KING

References and further reading

- Fine, K.** (1983) 'A Defence of Arbitrary Objects', *Proceedings of the Aristotelian Society*, supplementary vol. 57: 55-77. (An informal introduction to Fine's theory of arbitrary objects. The least technical of Fine's writings on the subject.)
- Fine, K.** (1985a) *Reasoning with Arbitrary Objects*, Oxford: Blackwell. (Fine's definitive and most technical work on arbitrary objects.)
- Fine, K.** (1985b) 'Natural Deduction and Arbitrary Objects', *Journal of Philosophical Logic* 14: 57-107. (A fairly technical work which presents some highlights of Fine (1985a).)
- Frege, G.** (1904) 'Was ist eine Funktion?', trans. 'What is a Function?', in *Translations from the Philosophical Writings of Gottlob Frege*, ed. P.T. Geach and M. Black, Oxford: Blackwell, 1977. (An attack on the view that instancial terms name indefinite objects.)
- Hazen, A.** (1987) 'Natural Deduction and Hilbert's ε -Operator', *Journal Of Philosophical Logic* 16: 411-21. (An outline of the view that instancial terms can be thought of as abbreviations of ε -terms. Informally written, but presupposes significant exposure to first-order logic and Fine (1985).)
- King, J.C.** (1987) 'Pronouns, Descriptions and the Semantics of Discourse', *Philosophical Studies* 51: 341-63. (Introduction to the view that instancial terms are contextually sensitive quantifier-like expressions of generality; very accessible.)
- King, J.C.** (1991) 'Instancial Terms, Anaphora and Arbitrary Objects', *Philosophical Studies* 61: 239-65. (Compares the account of instancial terms as context-dependent quantifiers to the account of Fine (1983, 1985). The presentation is informal; technical material in the appendix.)
- Leisenring, A.C.** (1969) *Mathematical Logic and Hilbert's ε -Symbol*, New York: Gordon & Breach. (The best English introduction to formal systems containing ε -terms; technical, but self-contained.)
- Quine, W.V.** (1950) *Methods of Logic*, New York: Holt; repr., Cambridge, MA: Harvard University Press, 4th edn, 1982. (System of natural deduction discussed in both Hazen (1987) and Fine (1985). The 1950 and 1982 editions differ significantly.)
- Wilson, G.** (1984) 'Pronouns and Pronominal Descriptions: A New Semantical Category', *Philosophical Studies* 45: 1-30. (The first paper to suggest that instancial terms are quantifier-like expressions of generality; very accessible.)

Quantifiers

The quantifiers ‘some’ and ‘every’ were the object of the very first logical theory, Aristotelian syllogistic. An example of a syllogism is ‘Every Spartan is Greek, every Greek is European, therefore every Spartan is European’. In such inferences, no quantifier is governed by another one. Contrast this with ‘Everybody loves somebody’. Modern logic is often taken to have begun when Frege systematized for the first time the logic of quantifiers, including such dependent ones. In general, much of what has passed as logic over the centuries is in effect the study of quantifiers. This is especially clear with the area of logic variously known as quantification theory, lower predicate calculus or elementary logic. Some philosophers have even sought to limit the scope of logic to such a study of quantifiers. Yet the nature of quantifiers is a delicate matter which is captured incompletely by the logic initiated by Frege and Russell.

1 Interpretation of the quantifiers

Traditionally, quantifiers have been taken to be those expressions which somehow indicate the size of the extension of a predicate (see **Logic, ancient §§1-3**; **Predicate calculus**), for example, expressions such as ‘some’ (the existential quantifier), ‘every’ (the universal quantifier), the numerical quantifiers ‘one’, ‘two’, ‘three’,..., ‘many’, ‘few’ and so on. The logical properties of the existential and universal quantifiers are studied in quantification theory or first-order logic. The numerical quantifiers are definable in terms of the existential and universal quantifiers, unlike quantifiers such as ‘many’ and ‘few’. For this reason the latter are sometimes called nonstandard quantifiers. Here we shall be concerned mainly with the standard quantifiers ‘some’ and ‘every’. When such quantifiers are applied to higher-order entities (predicates, relations, functions, predicates, relations and functions of predicates, relations and functions and so on), we are dealing with higher-order quantifiers (see **Second- and higher-order logics**).

The formulas of a first-order language L are built out of a set of nonlogical constants and individual variables by means of propositional connectives, the identity symbol and quantifiers. Typically, the formation rules for quantifiers allow us to take a formula with a free variable x and to prefix to it the existential quantifier ‘ $\exists x$ ’ or the universal quantifier ‘ $\forall x$ ’. The formula is then called the ‘scope’ of the quantifier in question (see **Scope**). In the resulting formula - for example, $(\exists x)(\dots x \dots)$ - the variable x is said to be ‘bound’ by the quantifier ‘ $\exists x$ ’. The nonlogical constants can be predicates, function symbols or individual constants.

A first-order formal language allows us to systematize the laws of quantification. Usually, a number of logically true (valid) formulas are assumed as logical axioms and a number of rules of inference are used to derive other logical truths (ideally, all of them) as theorems. A complete axiomatization of ordinary quantification theory was first given by Gödel in 1929. A theoretically more illuminating completeness proof was given by Henkin in 1949.

Quantification theory must be considered semantically, not only as a purely formal system. To specify the meaning of an interpreted first-order sentence is to specify its truth-conditions. Usually, such a specification of truth-conditions is done step by step, starting from inside, and at each step indicating how the truth or falsity of an expression depends on the truth or falsity of its component expressions. Such specifications of meaning are said to satisfy the principle of ‘compositionality’ (see **Compositionality**).

Any specification of truth-values must be relative to a model of the underlying language. The set of elements (individuals) of this model is called its domain. A quantifier is often said to range over this domain.

The truth-value of an interpreted and quantified sentence, say, $(\forall x)(\dots x \dots)$, cannot be specified in terms of the truth-value of the formula $(\dots x \dots)$, for the latter contains a free variable and hence does not have a truth-value. The truth-value of $(\forall x)(\dots x \dots)$ therefore has to be defined as a function of all the substitution instances of $(\dots x \dots)$ or else through the other semantic attributes of $(\dots x \dots)$. Tarski, who was the first to formulate an explicit truth-definition for quantificational languages (1933), chose the latter course and defined truth for first-order languages via the notion of satisfaction, so preserving the compositionality of his procedure (see **Tarski’s definition of truth**).

Tarski’s truth-definition is a formal specification of truth-conditions of sentences containing quantifiers, not yet a genuine interpretation of quantifiers. There are three main interpretations of quantifiers in the literature.

(A) *Quantifiers as higher-order predicates*. On this view, what, for example, the existential quantifier does in a sentence such as

(1) $(\exists x)S(x)$

is to say that the complex or simple predicate $S(x)$ is not empty. This is the approach we described above and is *prima facie* quite natural. It goes back to Frege. Its most important systematization and further development is the theory of so-called generalized quantifiers, currently the dominant trend in the study of quantifiers among theoretical linguists (see [Quantifiers, generalized](#)).

(B) *Substitutional interpretations of quantifiers*. On such an interpretation, the import of a sentence such as (1) is explained in terms of the set of different substitution-instances of the open formula $S(x)$ (see [Quantifiers, substitutional and objectual](#)).

(C) *Quantifiers as choice terms*. The basic idea can be explained by example. A natural interpretation of the sentence

(2) Every man owns a dog

is to say that for any man given to one, one can choose a dog such that the given man owns the chosen dog. In the formal counterpart of (2),

(3) $(\forall x)(\exists y)O(x, y)$,

we would say that, for every value of x , one can find a value of y such that $O(x, y)$. We can say that on this view quantifiers codify suitable choice terms. This way of interpreting quantifiers is systematized and generalized in Hintikka's game-theoretic semantics (see [Hintikka and Sandu 1996](#)).

2 The game-theoretic approach

In this game-theoretic treatment, a sentence such as (3) above is interpreted via a semantic game played by 'myself' (the initial verifier) against 'nature' (the initial falsifier), the former trying to produce true sentences while the latter is striving for the contrary. The universal quantifier ' $\forall x$ ' in (3) marks a move by 'nature', who must choose an individual, say d , from the universe of the model in which (3) is evaluated, and the existential quantifier ' $\exists y$ ' marks a move by 'myself', who must also choose an individual, say e . The sentence (3) is defined to be true if 'myself' has a method (usually codified by a mathematical function f) which, for every individual d chosen by 'nature' gives him the 'right' choice e , that is, $O(d, e)$ holds. More precisely, $O(d, f(d))$ holds (see [Semantics, game-theoretic](#)).

The central notion involved in interpretation (C) is that of a strategy, that is, a rule which specifies which move a player is to make in any possible situation which might come up in the game. The idea of an 'independent' quantifier is captured by that of an informationally independent choice. For instance, in (3) the choice of a value of y depends on the choice of a value of x in the sense that the choice connected with y is made on the basis of the player's knowledge of the value that was chosen for x , but not vice versa. In the jargon of game-theory, a game associated with a sentence of ordinary first-order language is one of perfect information.

In the usual formal (but interpreted) first-order languages, the three interpretations (A)-(C) assign the same truth-value to each sentence. Nevertheless, several things distinguish (C) from (A) and (B).

First, only interpretation (C) does justice to the fact that the expressive power of first-order languages lies in the dependence of quantifiers on each other. To understand first-order logic is, in this sense, to understand the idea of a dependent quantifier. In contrast, in (A) and (B), quantifiers are considered in isolation from each other. Interpretations of type (C) guided the work of David Hilbert and his school in the 1920s and 1930s see ([Hilbert and Bernays 1934](#)) and similar views have enjoyed a widespread albeit inconspicuous currency in mathematical practice.

Second, (C), unlike (A), yields a non-compositional interpretation of the standard quantifiers. An existential quantifier ' $\exists y$ ' is interpreted in (C) by a choice function whose arguments are certain universally quantified

variables. Which ones they are depends on the context.

Interpretation (C) offers an interesting possibility of generalizing quantification theory (first-order logic). If one understands the idea of a dependent quantifier as illustrated by (3), one *ipso facto* understands the idea of an independent quantifier. What often passes unnoticed is that the usual nested ordering of scopes for first-order logic rules out certain perfectly possible and understandable patterns of mutual dependence and independence between quantifiers, in that it is required that the scopes of two different quantifiers are either exclusive or nested, that is, that their scopes do not overlap only partially.

But what happens when this arbitrary restriction is given up? As an example, consider the first-order formula

$$(4) (\forall x)(\exists y)(\forall z)(\exists w)S(x, y, z, w).$$

The question is what happens if in (4) we require ‘ $\exists w$ ’ to be within the scope of ‘ $\forall z$ ’ but not in that of ‘ $\forall x$ ’? There is no way of expressing this complex of dependencies in the usual linear notation of first-order logic. For that purpose, a different notation has to be introduced. A special kind of informational independence has been studied by logicians under the title of ‘branching quantifiers’. They were introduced by Henkin in 1959. A more general way of indicating independence is to use a slash notation:

$$(5) (\forall x)(\exists y)(\forall z)(\exists w/\forall x)S(x, y, z, w).$$

The idea here is that the last existential quantifier does not depend on ‘ $\forall x$ ’.

What can be said of quantifiers can also be said, *mutatis mutandis*, of disjunctions and conjunctions: the independence indicator slash can be applied to them too. The languages using the slash notation are called independence-friendly first-order languages.

Neither approach (A) nor approach (B) works for independence-friendly first-order languages. The reason is that such a language involves context dependencies which make a compositional interpretation unrealistic. For a compositional interpretation requires that the meaning of an expression be specified in terms of its component parts. This is not the case with the expression $(\exists w/\forall x)S(x, y, z, w)$ in (5), for its interpretation makes reference to an element ‘ $\forall x$ ’ which occurs outside it.

In contrast, interpretation (C) extends naturally to cover this new language. All we need is to drop the requirement that the semantic game associated with a sentence such as (5) is one of perfect information.

What has been said has a number of consequences not usually noted in the study of quantifiers.

- (1) The meaning of quantifiers is often explained in terms of their ‘ranging over’ a set of entities. Such an explanation presupposes an interpretation along the lines of (A) or (B) and hence is not a general explanation of the meaning of quantifiers.
- (2) The notion of quantifier scope is ambiguous. The usual formal scope notation indicates two different things. On the one hand, it indicates the relative logical priority of different quantifiers (and connectives); on the other, it indicates the segment of a formula (or a discourse) in which a variable is bound by a given quantifier. There is no reason to expect that these two always go together, and in natural languages they seem to involve entirely different mechanisms.
- (3) Certain types of quantifiers, namely informationally independent ones, do not admit of a natural compositional truth-definition, but have to be treated in other ways semantically.
- (4) A general theory of quantification will have to comprise more than is currently known as quantification theory or ordinary first-order logic.
- (5) Since independence-friendly first-order logic is not axiomatizable, it is unrealistic to expect the logic of quantifiers to admit of a complete systematization.
- (6) The logical behaviour of quantifiers in natural languages seems to differ more radically from ordinary formal first-order logic than has usually been recognized.

See also: [Logical constants](#); [Logical and mathematical terms, glossary of](#); [Quantification and inference](#)

JAAKKO HINTIKKA
GABRIEL SANDU

References and further reading

- Frege, G.** (1879) *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*, Halle: Nebert; trans. J. van Heijenoort, 'Begriffsschrift, a Formula Language, Modelled Upon That of Arithmetic, for Pure Thought', in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic 1879-1931*, Cambridge, MA: Harvard University Press, 1967. (First explicit logic of quantifiers.)
- Gödel, K.** (1929) 'Über die Vollständigkeit des Logikkalküls' (doctoral dissertation, University of Vienna), in *Kurt Gödel: Collected Works*, vol. 1, *Publications 1929-1936*, ed. S. Feferman, J.W. Dawson Jr, S.C. Kleene, G.H. Moore, R.M. Solovay and J. van Heijenoort, Oxford: Oxford University Press, 1986, 60-101. (First completeness proof for ordinary first-order logic of quantifiers.)
- Goldfarb, W.** (1979) 'Logic in the Twenties: The Nature of the Quantifier', *Journal of Symbolic Logic* 44: 351-68. (Changing conceptions of quantifiers from Frege to Hilbert.)
- Henkin, L.** (1949) 'Completeness in the Theory of Types', *Journal of Symbolic Logic* 14: 159-66. (Greatly improved completeness proof.)
- Henkin, L.** (1959) 'Some Remarks on Infinitely Long Formulas', in *Infinitistic Methods*, Oxford: Pergamon, 167-83. (First use of informationally independent quantifiers.)
- Hilbert, D. and Bernays, P.** (1934) *Grundlagen der Mathematik I-II*, Berlin: Springer. (Quantifiers and choice functions.)
- Hintikka, J. and Sandu, G.** (1996) 'Game-Theoretical Semantics', in J. van Benthem and A. ter Meulen (eds) *Handbook of Logic and Language*, Amsterdam: Elsevier. (A survey of game-theoretic semantics.)
- Hintikka, J. and Kulas, J.** (1985) *Anaphora and Definite Descriptions*, Dordrecht: Reidel. (Quantifiers in natural languages.)
- Kripke, S.A.** (1976) 'Is There a Problem about Substitutional Quantification?', in G. Evans and J. McDowell (eds) *Truth and Meaning*, Oxford: Oxford University Press, 325-419. (A spirited defence of the substitutional interpretation.)
- Tarski, A.** (1933) *Pojęcie prawdy w językach nauk dedukcyjnych*, Warsaw; trans. J.H. Woodger (1956), 'On the Concept of Truth in Formalized Languages', in *Logic, Semantics, Metamathematics*, ed. J. Corcoran, Indianapolis, IN: Hackett Publishing Company, 2nd edn, 1983, 152-278. (The classic work on the concept of truth.)
- Westerståhl, D.** (1989) 'Quantifiers in Formal and Natural Languages', in D. Gabbay and F. Guenther (eds) *Handbook of Philosophical Logic*, Dordrecht: Reidel, vol. 4, 1-131. (A survey of the theory of generalized quantifiers.)

Quantifiers, generalized

Generalized quantifiers are logical tools with a wide range of uses. As the term indicates, they generalize the ordinary universal and existential quantifiers from first-order logic, ‘ $\forall x$ ’ and ‘ $\exists x$ ’, which apply to a formula $A(x)$, binding its free occurrences of x . $\forall x A(x)$ says that $A(x)$ holds for all objects in the universe and $\exists x A(x)$ says that $A(x)$ holds for some objects in the universe, that is, in each case, that a certain condition on $A(x)$ is satisfied. It is natural then to consider other conditions, such as ‘for at least five’, ‘at most ten’, ‘infinitely many’ and ‘most’. So a quantifier Q stands for a condition on $A(x)$, or, more precisely, for a property of the set denoted by that formula, such as the property of being non-empty, being infinite, or containing more than half of the elements of the universe. The addition of such quantifiers to a logical language may increase its expressive power.

A further generalization allows Q to apply to more than one formula, so that, for example, $Qx(A(x), B(x))$ states that a relation holds between the sets denoted by $A(x)$ and $B(x)$, say, the relation of having the same number of elements, or of having a non-empty intersection. One also considers quantifiers binding more than one variable in a formula. $Qxy, zu(R(x, y), S(z, u))$ could express, for example, that the relation (denoted by) $R(x, y)$ contains twice as many pairs as $S(z, u)$, or that $R(x, y)$ and $S(z, u)$ are isomorphic graphs.

In general, then, a quantifier (the attribute ‘generalized’ is often dropped) is syntactically a variable-binding operator, which stands semantically for a relation between relations (on individuals), that is, a second-order relation. Quantifiers are studied in mathematical logic, and have also been applied in other areas, notably in the semantics of natural languages. This entry first presents some of the main logical facts about generalized quantifiers, and then explains their application to semantics.

1 The general concept

A (generalized) quantifier is a relation between relations on individuals (in other words, a second-order relation). Its type, a finite sequence of positive numbers, indicates how many relations and of what degree. For example, a quantifier of type $\langle 2, 1, 1, 4 \rangle$ is a relation between one binary relation (that is, a relation of order 2), two sets and one four-place relation. Let M be a non-empty universe. Then define a (local) quantifier of type $\langle 2, 1, 1, 4 \rangle$ on M to be any relation between a binary relation on M , two subsets of M and a four-place relation on M . Then a (global) quantifier Q of type $\langle 2, 1, 1, 4 \rangle$ is a function which associates with each universe M a quantifier Q_M of the same type on M .

The quantifiers ‘ $\forall$ ’ and ‘ $\exists$ ’ are thus of type $\langle 1 \rangle$. Below are definitions of these and some other type $\langle 1 \rangle$ quantifiers, with M any universe, A any subset of M , and $|A|$ the cardinality of A (Q^R is called the Rescher quantifier):

$$\forall_M A \Leftrightarrow A = M$$

$$\exists_M A \Leftrightarrow A$$

$$(\exists_{\geq 5})_M A \Leftrightarrow |A| \geq 5$$

$$(\exists_{\leq 10})_M A \Leftrightarrow |A| \leq 10$$

$$(Q_0)_M A \Leftrightarrow A$$

$$(Q^R)_M A \Leftrightarrow |A| > |M - A|$$

Here are some more examples of quantifiers (types are written to the left):

$$\langle 1, 1 \rangle \quad I_M AB \Leftrightarrow |A| = |B| \quad (\text{the H\"artig quantifier})$$

$$\langle 1, 1, 1 \rangle \quad \text{more...than}_M ABC \Leftrightarrow |A \cap C| > |B \cap C|$$

$$\langle 2 \rangle \quad W_M R \Leftrightarrow R \text{ is a well-order of } M$$

$$\langle 2, 2 \rangle \quad IG_M RS \Leftrightarrow R \text{ and } S \text{ are isomorphic graphs}$$

The notion of a type $\langle 1 \rangle$ quantifier was introduced by Andrzej Mostowski (1957), and extended to arbitrary types by Per Lindström (1966). Aristotle’s syllogisms (see [Logic, ancient §§1-3](#)) involve type $\langle 1, 1 \rangle$ quantifiers, a natural class, as will be seen in §3. The general notion of quantifiers as variable-binding operators denoting

second-order relations was in fact explicit in Frege's work, except that Frege did not relativize to an arbitrary universe (see Frege, G. §6).

The quantifiers above share an important property, closure under isomorphism (or ISOM). We use the following notation: a function f from a set X to a set Y is 'lifted' to subsets A of X , or to (for example) ternary relations R on X , by $f(A) = \{f(a) : a \in A\}$ and $f(R) = \{(f(a), f(b), f(c)) : (a, b, c) \in R\}$. Now for a quantifier Q of type $\langle 2, 1, 1, 4 \rangle$, say, define Q to be *closed under isomorphism* if whenever f is a bijection from M to M' (a one-one function from M onto M'), and R, A, B, S are appropriate arguments for Q_M ,

$$Q_M R A B S \Leftrightarrow Q_{M'} f(R) f(A) f(B) f(S).$$

ISOM says that Q does not care about the individual elements of a universe, only about their 'structure' with respect to the relational arguments. This allows quantifiers to be treated as logical constants (see §2 below; Logical constants §3). ISOM has a particularly clear meaning for quantifiers which relate sets, so-called 'monadic' quantifiers (that is, of type $\langle 1, 1, \dots, 1 \rangle$). It then says that Q only cares about the cardinalities of these sets (and sets that can be formed from them by taking unions, intersections and complements), not the sets themselves. For example, for a type $\langle 1, 1 \rangle$ quantifier, ISOM is equivalent to the following condition:

If $A, B \subseteq M$, $A', B' \subseteq M'$, $|A \cap B| = |A' \cap B'|$, $|A - B| = |A' - B'|$, $|B - A| = |B' - A'|$, and $|M - (A \cup B)| = |M' - (A' \cup B')|$, then $Q_M A B \Leftrightarrow Q_{M'} A' B'$.

Thus, in the monadic case, ISOM quantifiers deal with *quantities*, as the name suggests they should. Indeed, such a quantifier can be redefined as a relation between numbers; the examples above provide illustrations.

2 Quantifiers in logic

Generalized quantifiers are easy to add as new logical constants to first-order logic, FO. For example, 'A is finite' cannot be expressed in FO, but if we add the quantifier Q_0 above, the sentence $\neg Q_0 x P x$ expresses precisely that the denotation of P is finite. If Q is a quantifier of type $\langle 2, 1, 1 \rangle$, say, the logic $\text{FO}(Q)$ is formed from FO by adding a formation rule allowing formulas of the form

$$(1) Qxy, z, u(A(x, y), B(z), C(u)),$$

where $A(x, y)$, $B(z)$ and $C(u)$ are formulas. Free occurrences of the indicated variables become bound in (1). To interpret (1) in a model M with universe M , note that relative to M and the variables x, y (and a variable assignment which we suppress here), $A(x, y)$ denotes a binary relation $A(x, y)^M$ on M , namely, the set of pairs (a, b) such that $A(a, b)$ holds in M . Then, (1) is defined to be true in M if and only if $Q_M A(x, y)^M B(z)^M C(u)^M$. This generalizes the usual clauses for $\forall x A(x)$ and $\exists x A(x)$ in the truth-definition for FO. Logical consequence in $\text{FO}(Q)$ is defined as usual (the conclusion must be true in every model of the premises - see Model theory). And just as FO can be used to formulate theories, like arithmetic or group theory (first-order theories), so can $\text{FO}(Q)$. The properties of these theories, and of the consequence relation, will of course depend on Q .

Similarly, one defines logics of the form $\text{FO}(Q_1, \dots, Q_n)$. This is a model-theoretic conception of a logic as given roughly by a class of sentences, a class of models and a truth-relation between them (in contrast, a proof-theoretic conception would emphasize a formal inference system). Some basic constraints apply; for example, only ISOM quantifiers are considered.

Abstract model theory studies the properties of model-theoretic logics in general, logics with generalized quantifiers being prime examples. Among interesting properties of a logic L are axiomatizability (existence of a formal inference system proving precisely the valid sentences of L), compactness (if every finite subset of a set Φ of L -sentences has a model then the whole of Φ has a model) and the Löwenheim property (if an L -sentence has an infinite model it has a denumerable model; see Löwenheim-Skolem theorems and nonstandard models §§2, 3). FO has these properties and several stronger properties as well, and a celebrated result of abstract model theory is that such properties in fact characterize FO. To state this one needs a way to compare the expressive power of various logics. A logic L' is said to be an extension of a logic L , $L \leq L'$, if everything that can be said in L can also be said in L' , in other words if for every L -sentence there is a logically equivalent L' -sentence (one with the

same models). Also, $L \equiv L'$ if $L \leq L'$ and $L' \leq L$. Note that if L is a logic with generalized quantifiers, $\text{FO} \leq L$. Given a simple assumption about these quantifiers, one can prove the following result (Lindström's theorem):

If L is either compact or axiomatizable, and has the Löwenheim property, then $\text{FO} \equiv L$.

Thus, increase of expressive power entails loss of some of FO's properties. For example, if 'A is finite' is expressible in L , it can be seen that L can be neither compact nor axiomatizable. But there do exist compact and axiomatizable proper extensions of FO (hence the Löwenheim property fails for them), the simplest example being $\text{FO}(Q_1)$, where Q_1 is the quantifier 'there exist uncountably many'. Extending FO should be done with care, to strike a balance between expressive power and model-theoretic behaviour.

How many different quantifiers (of a given type) are there? Locally, the rough answer is: many times more than the size of the universe (see [Westerståhl 1989](#): §4.6 for more detail). For example, if M has n elements, then there are 2^{4^n} type $\langle 1, 1 \rangle$ quantifiers on M . So there are $2^{16} = 65,536$ type $\langle 1, 1 \rangle$ quantifiers on a universe with 2 elements! Counting only ISOM quantifiers the number is 1,024, still large compared to 2. As to global quantifiers, even if we restrict attention to finite universes there are uncountably many ISOM quantifiers of each type.

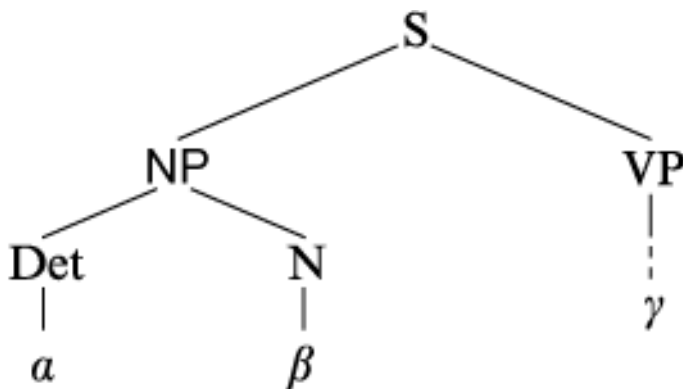
Some of these different quantifiers will have the same expressive power. For example, $\exists_{\geq 5}$ is definable in FO, so $\text{FO}(\exists_{\geq 5}) \equiv \text{FO}$. In general, a quantifier Q is said to be definable in a logic L if it is defined by some L -sentence whose nonlogical symbols match the type of Q . So for a Q of type $\langle 2, 1, 1 \rangle$ to be definable in L there must exist an L -sentence with nonlogical symbols P_1 , P_2 , and P_3 , with P_1 binary and P_2 , P_3 unary, which is logically equivalent to $Qxy, z, u(P_1xy, P_2z, P_3u)$. We have $\text{FO}(Q_1, \dots, Q_n) \leq L$ if and only if each Q_i is definable in L .

The number of non-equivalent logics of the form $\text{FO}(Q)$ is still uncountable, for each type of Q . Moreover, the types yield a strict hierarchy of expressive power. Types can be compared with respect to the largest degree of the relation arguments. For example, $\langle 2, 1, 1, 3, 3 \rangle$ is lower than $\langle 2, 4 \rangle$. A more fine-grained (partial) order also takes into account the number of arguments of each degree. Then $\langle 3, 1, 2, 3 \rangle$ is lower than $\langle 2, 1, 1, 3, 3 \rangle$ since it has fewer unary arguments but the same number of binary and ternary arguments. It can be proved that in each type there are quantifiers not definable in any logic $\text{FO}(Q_1, \dots, Q_n)$ with $Q_1, \dots, Q_n$ of lower types (in the more fine-grained order). Again this holds even if only finite models are considered.

The study of quantifiers in mathematical logic has either been in the direction of abstract model theory or focused on logics with particular quantifiers (and theories formulated in such logics), a prominent case being $\text{FO}(Q_1)$ and certain of its extensions. Quantifiers have also found applications outside mathematics. Restricted to finite models, they have turned up in computer science, in attempts to find extensions of FO 'capturing' certain complexity classes from the theory of computational complexity. They have been applied to modal logic, and to the study of conditionals. But perhaps the most striking application concerns the semantics of natural languages, to which we now turn.

3 Quantifiers and natural language

Consider the following sentence form, common to many languages.



The sentence (S) consists of a noun phrase (NP) and a verb phrase (VP), and the NP in turn consists of a determiner (Det) and a noun (N). A simple English example is

(2) No student smokes

but Det, N and VP could be complex expressions. Now, given a universe M , N typically denotes a subset A of M , and VP another subset, B . Thus, Det is most straightforwardly taken to denote a binary relation between subsets of M , that is, a type $\langle 1, 1 \rangle$ quantifier on M . Varying M , Det denotes a global type $\langle 1, 1 \rangle$ quantifier.

In (2), this quantifier is $no_M AB \Leftrightarrow A \cap B = \emptyset$. Other Dets denote other quantifiers:

$$all_M AB \Leftrightarrow A \subseteq B$$

$$at\ least\ five_M AB \Leftrightarrow |A \cap B| \geq 5$$

$$most_M AB \Leftrightarrow |A \cap B| > |A - B|$$

$$all\ but\ three_M AB \Leftrightarrow |A - B| = 3$$

$$more\ than\ two-thirds\ of\ the_M AB \Leftrightarrow |A \subseteq B| > \frac{2}{3}|A|$$

$$all \dots except\ John_M AB \Leftrightarrow A - B = \{j\}$$

The last example comes from a ‘discontinuous’ Det, as in, for example, ‘All students except John smoke’. A few Dets take more than one noun argument, as in

More students than teachers smoke,

which involves the type $\langle 1, 1, 1 \rangle$ quantifier ‘more...than’ mentioned in §1.

Furthermore, NPs naturally denote type $\langle 1 \rangle$ quantifiers: ‘at least five students’ denotes the set of subsets of M containing at least five students. In general one obtains such NP denotations from Det denotations by fixing a particular noun. Also NPs without Dets, such as ‘John’, or ‘Bill and Sue’, may be taken to denote type $\langle 1 \rangle$ quantifiers:

$$John_M B \Leftrightarrow j \in B$$

$$Bill\ and\ Sue_M B \Leftrightarrow b \in B \ \& \ s \in B$$

This treatment of proper names may seem complex, but has the advantage of allowing all NPs to be interpreted in the same way.

Thus, generalized quantifiers appear in a linguistic context not by means of some technical trick or invention, but rather as the most straightforward way to treat the semantics of noun phrases.

Still, is it really necessary to bring in generalized quantifiers here? After all, ‘all’ can be defined in FO by $\forall x(Ax \rightarrow Bx)$ and, similarly, quantifiers such as ‘at least five’. Nevertheless, sticking to FO is not advisable. First, the use of type $\langle 1, 1 \rangle$ quantifiers respects the syntactic form of sentences whereas a translation into FO does not. So generalized quantifiers facilitate a compositional semantics for natural language (see **Compositionality**). Second, in important cases a translation into FO is simply impossible. For example, proportional quantifiers such as ‘most’ or ‘more than two thirds of the’ are undefinable in FO. That is, there is no sentence in first-order logic, however complex, which adequately accounts for the truth-conditions of, say, ‘Most students smoke’. In fact, one can even prove that there is no such sentence in any logic FO($Q_1, \dots, Q_n$) with $Q_1, \dots, Q_n$ of type $\langle 1 \rangle$. So type $\langle 1, 1 \rangle$ quantifiers are crucial also with respect to expressive power.

Most Dets have ISOM denotations, with putative exceptions including ‘every...but John’ or ‘John’s’. If ‘John’ denotes a fixed individual ISOM will fail, but there are various strategies for avoiding such Dets. The salient property of Det denotations, however, is the special role played by the first argument, coming from the noun: it restricts the domain of quantification. Formally this is reflected in two constraints, called ‘conservativity’ and ‘extension’: for all M , $A, B \subseteq M$ and $M \subseteq M'$,

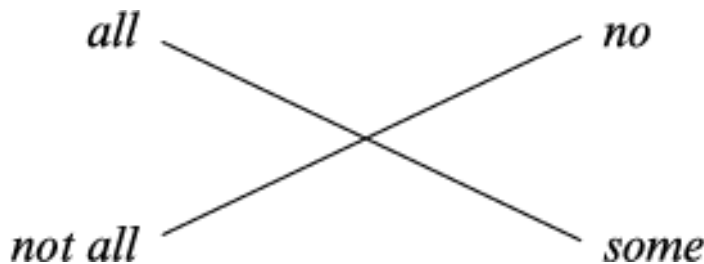
$$\text{CONS} \quad Q_M AB \Leftrightarrow Q_M A \ A \cap B$$

$$\text{EXT} \quad Q_M AB \Leftrightarrow Q_{M'} AB$$

CONS says that only that part of the verb argument matters which is common to the noun argument. Indeed, ‘All but three students smoke’ is equivalent to ‘All but three students are students which smoke’. CONS predicts, for example, that the Härtig quantifier $rmI_M AB \Leftrightarrow |A| = |B|$ (for which CONS fails) is not the denotation of any Det. EXT implies that the rest of the universe - the part lying outside the noun denotation - is irrelevant. It predicts, for example, that there can be no Det meaning ‘some’ on universes with, say, fewer than ten elements, and meaning ‘most’ on other universes.

Discounting a few marginal (and debatable) exceptions, all Det denotations appear to satisfy CONS and EXT. Other properties single out subclasses of quantifiers. Call Q ‘intersective’ if only the intersection of the noun and the verb argument matters; $A \cap B = A' \cap B'$ implies $Q_M AB \Leftrightarrow Q_M A'B'$. Examples are ‘some’, ‘at least five’, ‘exactly three’ and ‘only finitely many’. This turns out to be related to the acceptability of the corresponding Dets in ‘there’-constructions: ‘There are some/at least five/no cows in the barn’ is fine, but ‘There are all but five/most/all cows in the barn’, with non-intersective Dets, is not.

Monotonicity properties are also significant. Q is ‘right monotone increasing’ (or decreasing) if $Q_M AB$ and $B \subseteq B'$ (or $B' \subseteq B$) implies $Q_M AB'$. Similarly for monotonicity in the left argument. Many Dets denote right monotone quantifiers. For example, ‘more than two thirds of the’ is right monotone increasing, and ‘at most one third of the’ is right monotone decreasing (neither is left monotone). And many non-monotone cases, such as ‘(exactly) five’, ‘between three and five’ and ‘ten per cent of’, are conjunctions of a decreasing and an increasing quantifier. Right monotonicity has been related to various linguistic phenomena, in particular so-called ‘negative polarity items’. Left monotonicity is more scarce among Det denotations. It may be shown that, over finite models and assuming CONS and EXT, all left monotone quantifiers are first-order definable. Left monotone Det denotations are usually also right monotone. The four possible kinds of ‘double monotonicity’ are exemplified in the Aristotelian square of opposition:



Indeed, given CONS and a certain uniformity requirement, it can be proved that these are the *only* doubly monotone quantifiers (without this requirement there are other examples: ‘at most ten’, ‘all but at most three’, ‘infinitely many’).

NP and Det denotations are normally monadic: they relate sets. But sentences can combine several NPs with a transitive verb:

(3) Most students know at least three teachers.

The truth-conditions of (3) are an *iteration* of the two type $\langle 1, 1 \rangle$ quantifiers involved: formally (with A the set of students, B the set of teachers and R the relation of knowing), $most_M A \{a \text{ at least } three_M B \{b \text{ } Rab\}\}$. However, other seemingly similar sentences combine the quantifiers differently, for example:

(4) 1,200 philosophers wrote 2,000 entries for the Encyclopedia,

which does *not* mean that each of the 1,200 philosophers wrote 2,000 entries, as an iterated reading would have it, but rather that each of them wrote *some* entry, and that each of the 2,000 entries was written by *some* one (or more) of them. This mode of combination has been called ‘cumulation’. Another suggested mode is ‘branching’:

(5) Most of the students and most of the teachers all know each other

appears to have the reading that there is a large subset X of the set of students (containing more than half of them) and a large subset Y of the set of teachers such that every student in X and every teacher in Y know each other. Again this is not an iterated reading. But note that (3), (4) and (5) can be accounted for by quantifiers of type

$\langle 1, 1, 2 \rangle$, taking two set arguments from the nouns and one relation argument from the verb. For (3) and (4) this quantifier is definable from the monadic Det denotations, but not so for (5). More precisely, it can be shown that the branching of ‘most’ as in (5) is not definable in FO/(‘most’). Examples like this indicate a possible use also for polyadic (non-monadic) quantifiers in the semantics of natural languages.

See also: [Logical and mathematical terms, glossary of](#)

DAG WESTERSTÅHL

References and further reading

- Barwise, J. and Feferman, S.** (eds) (1985) *Model-Theoretic Logics*, Berlin: Springer. (Classical, handbook-style reference for abstract model theory and the mathematics of quantifiers.)
- Benthem, J. van** (1986) *Essays in Logical Semantics*, Dordrecht: Reidel. (Includes some of the papers that initiated the subject of generalized quantifiers and natural language.)
- Benthem, J. van and Westerståhl, D.** (1995) ‘Directions in Generalized Quantifier Theory’, *Studia Logica* 55: 389-419. (A quick survey of work on generalized quantifiers at the time of writing.)
- Keenan, E.L. and Westerståhl, D.** (1997) ‘Generalized Quantifiers in Linguistics and Logic’, in J. van Benthem and A. ter Meulen (eds) *Handbook of Logic and Language*, Amsterdam: Elsevier, 837-93. (Updates the exposition in Westerståhl (1989), from a more linguistic perspective.)
- Krynicky, M., Mostowski, M. and Szczerba, L.W.** (eds) (1995) *Quantifiers: Logics, Models and Computation*, vol. 1, *Surveys*, Dordrecht: Kluwer. (Surveys mathematical work on quantifiers between the time of appearance of Barwise and Feferman (1985) and 1992.)
- Lindström, P.** (1966) ‘First-Order Predicate Logic with Generalized Quantifiers’, *Theoria* 32: 186-95. (The extension to arbitrary types of the notion of a type $\langle 1 \rangle$ quantifier.)
- Mostowski, A.** (1957) ‘On a Generalization of Quantifiers’, *Fundamenta Mathematicae* 44: 12-36. (The introduction of the notion of a type $\langle 1 \rangle$ quantifier.)
- Westerståhl, D.** (1989) ‘Quantifiers in Formal and Natural Languages’, in D. Gabbay and F. Guenther (eds) *Handbook of Philosophical Logic*, vol. 4, Dordrecht: Reidel, 1-131. (Detailed exposition of much of the material in this entry.)

Quantifiers, substitutional and objectual

Understood substitutionally, 'Something is F' is true provided one of its substitution instances (a sentence of the form 'a is F') is true. This contrasts with the objectual understanding, on which it is true provided 'is F' is true of some object in the domain of the quantifier. Substitutional quantifications have quite different truth-conditions from objectual ones. For instance, 'Something is a mythological animal' is true if understood substitutionally, since the substitution instance 'Pegasus is a mythological animal' is true. But understood objectually, the sentence is not true, since there are no mythological creatures to make up a domain for the quantifier.

Since substitutional quantifiers do not need domains over which they range, it is easy to introduce substitutional quantifiers which bind predicate or sentential variables, even variables within quotation marks. One reason for interest in substitutional quantification is the hope that it may provide a way to understand discourse which appears to be about numbers, properties, propositions and other 'troublesome' sorts of entities as being free of exceptional ontological commitments. Whether natural language quantification is sometimes plausibly construed as substitutional is not, however, clear.

1 Objectual and substitutional quantification

The existential quantification

(1) $\exists x$ x is a logician

is true if 'x is a logician' is true in at least one case; the universal quantification

(2) $\forall x$ x is a logician

is true if 'x is a logician' is true in every case. Two accounts may be given of what it is for 'x is a logician' to be true for some or all cases. These yield two kinds of quantification: 'objectual' and 'substitutional'.

In an objectual account, the quantifier has a domain of objects; sentences such as 'x is a logician' are related to the contents of the quantifier's domain by the 'true of' relation. The sentence 'x is a logician' is true of a member of the domain if and only if that member is a logician; 'x taught y' is true of a pair $\langle u, v \rangle$ drawn from the domain just in case u taught v , and so on. Understood objectually, (1) is true if 'x is a logician' is true of at least one thing in the domain; (2) is true if 'x is a logician' is true of everything in the domain.

An objectual quantifier is tied directly to the domain of objects over which it quantifies. A substitutional quantifier does not have a domain; it is not a device whose semantics is explained in terms of the 'true of' relation. Instead, a substitutional quantifier has a 'substitution class', a collection of expressions which may be substituted for the variables the substitutional quantifier binds. The 'substitution instances' of a quantified sentence such as (1) or (2) are the results of removing the initial quantifier and replacing what is bound with some member of the substitution class. For instance, if the substitution class consists of the names 'Frege' and 'Russell' (the names, not the people), the substitution instances of (1) and (2) are the sentences

Frege is a logician.

Russell is a logician.

In substitutional quantification the cases to be considered are substitution instances. Read substitutionally, (1) is true provided at least one of its substitution instances is true; (2) is true provided they all are.

In order to avoid confusion, it is best to use distinct notations for the two kinds of quantifiers. Hereafter ' $\exists$ ' and ' $\forall$ ' are the objectual quantifiers, ' x ' and ' y ' their variables; ' Σ ' is the substitutional existential, ' Π ' the substitutional universal quantifier, and ' X ' and ' S ' their variables.

How different are $\exists x \phi x$ and $\Sigma X \phi X$? Suppose the domain of ' $\exists$ ' contains things not named by anything in the substitution class of ' Σ '. For example, say the domain of ' $\exists$ ' is the set of all animals, and let the substitution class for ' Σ ' be the proper names of animals. Not every horse has a name (proper or otherwise), so ' $\exists x$ x is a horse and x has no name' is true. But ' ΣX X is a horse and X has no name' is false. If it were true, then

α is a horse and α has no name,

where α is a proper name of a horse, would be true. But if α names something, then ' α has no name' is clearly false.

Another difference arises because there can (apparently) be true sentences containing empty (non-referring) names, such as 'Pegasus is only a mythological animal'. If the animal name 'Pegasus' does not name anything, *a fortiori* it does not name anything in the domain of an objectual quantifier. So, if the domain of ' $\exists$ ' is all animals, the sentence ' $\exists x x$ is only a mythological animal' is false: there is not any animal which is only mythological. But the sentence ' $\Sigma X X$ is only a mythological animal' is true, as one of its substitution instances is true.

The quantifiers of (1) and (2) are 'nominal' quantifiers - their variables occupy only the positions of names or singular terms. To further appreciate the difference between substitutional and objectual quantifiers, let us consider how we might introduce *non-nominal* quantifiers into a language. To introduce a substitutional quantifier whose variables can occupy, say, the positions of predicates or whole sentences, all one needs to do is to specify a substitution class, add new variables and define 'substitution instance'. For example, if we take expressions such as 'is Greek' to comprise the substitution class, we can introduce a substitutional quantifier which, like that of (1') ΣX (Socrates X and Plato X),

binds predicate variables. (Examples of substitution instances thereof would be 'Socrates is wise and Plato is wise' and 'Socrates snores and Plato snores'.) If sentences make up the substitution class, we can write, for example,

(2') ΠS (Dick said that $S \rightarrow$ it is false that S),

with a quantifier binding sentential variables. The truth-conditions of these sentences are exactly as before: (1') is true provided at least one substitution instance is true; (2') is true provided all are. Provided all the instances already had truth-values, such substitutional quantifications automatically have truth-values.

To introduce objectual quantifiers whose variables can occupy the positions of predicates or sentences requires one to assign domains to such quantifiers and to explain what it is for something such as 'Socrates X ' or 'Dick said that X ' to be true of something in the domain of the quantifier. One can do this only if there are objects to make up the domain. Since appropriate domains for predicate and sentential variables seem to be, respectively, sets (or, perhaps, universals) and propositions, someone with doubts about their existence ought to have doubts about the objectual analogues of (1') and (2') (see [Nominalism §4](#)). No such worries attach to (1') and (2') themselves.

Some have hoped that substitutional quantification would provide an innocuous means of interpreting non-nominal natural language quantifications, such as 'Bill did all that John did' and 'There's something Socrates is that Plato is too'. Some hope it will interpret apparently nominal quantifications ('Whatever Dick said is false') and numerical quantification, whose domains seem to some ontologically suspect.

Another reason for interest in substitutional quantification is worries about the coherence of using an objectual quantifier to 'quantify into' positions like that occupied by 'Jane' in

It is possible that Jane is happy.

Nancy thinks that Jane is happy.

The position of 'Jane' in these sentences is 'opaque' (it is a position for which substitution of terms which pick out the same thing can lead from truth to falsehood; see [Propositional attitude statements §2](#); [Modal logic, philosophical issues in §3](#)). Quine has argued repeatedly that an objectual quantifier cannot bind a variable in opaque position (see [Quine 1951](#)). Now, many perfectly good English sentences, for example,

(3) Some unhappy woman could have been happy.

There's someone [such that] Nancy thinks [that she] is happy.

seem to involve quantification into opaque position. Since there is no problem with a substitutional quantifier binding a variable in opaque position, one might conclude that sentences such as in (3) involve substitutional quantification.

In the author's opinion, it is implausible that English quantification into opaque positions as in (3) can or should be interpreted substitutionally. For one thing, Quine's arguments that an objectual quantifier cannot bind a variable in opaque position are unsound (see [Kaplan 1986](#)). Furthermore, the way we mix quantifiers and expressions such as 'believes' prohibits a substitutional reading of the quantifier. Consider

(4) Twain is Clemens; Sal believes that Twain is sad; Sal does not believe that Clemens is sad.

This does not imply

(5) Twain is something such that Sal believes that it is sad and Twain is something such that Sal does not believe that it is sad,

since (4) could be true, but (5) ascribes contradictory properties to Twain. Now (5) seems to have the form

(6) $(\exists x)(x = \text{Twain} \ \& \ \text{Sal believes that } x \text{ is sad}) \ \& \ (\exists x)(x = \text{Twain} \ \& \ \text{Sal does not believe that } x \text{ is sad}).$

If the existential quantifier 'E' in (6) is objectual, (4) does not imply (6), since no object in the domain could make both conjuncts true. But if it is substitutional, (4) does imply (6), and hence (5). But we have already said that (4) does not imply (5). So the quantifier is not substitutional. (Incidentally, this shows that when a substitutional quantifier binds variables in opaque positions, $\exists x\phi x$ and $\Sigma X\phi X$ may have different truth-values even if the substitution class for ' Σ ' consists only of names of objects in the domain of ' $\exists$ ', and every such object is named by a member of the class.)

The above is not meant to suggest that one cannot substitutionally quantify into opaque positions, but only that English quantification into such positions is not substitutional. In fact, substitutional quantifiers are very useful when one wants to bind variables in positions, such as those inside quotation marks, where it is indeed impossible in English to bind a referential variable. A formula along the lines of

(7) $\Pi S: 'S' \text{ is true if and only if } S$

is perfectly intelligible, and may even, in certain cases, be looked upon as 'defining truth' (see [Kripke 1976](#) and [Grover 1992](#) for discussion).

2 Interpreting natural language quantifiers

Natural language quantifiers cannot in general be plausibly understood as substitutional. Is quantification in English (or other natural languages; the disjunct is dropped hereafter) ever plausibly taken to be substitutional? If so, what, if anything, might this say about the 'ontological commitments' of speakers?

Some philosophers ([Gottlieb](#), for example) have suggested that English quantifiers are ambiguous, between substitutional and objectual ones. Others ([Marcus 1993](#), for example) have suggested that substitutional quantifiers provide, in some sense, the primary interpretation of English idioms of quantification, the objectual interpretation being a limiting or special case.

Arguments for the ambiguity thesis often turn on intuitions about what (the use of) a sentence ontologically commits one to. Roughly speaking, a sentence S ontologically commits one to there being an F provided that, as a matter of logic, one cannot consistently hold that S is true but there is not an F . (More precisely, let F be a singular English noun and let s name the English sentence S . s ontologically commits one to there being an F if the sentence 'If S , then there is an F ' is logically true, when 'there is' is understood objectually. To say ' s commits one to a ' is to say that ' a exists' follows from S . See [Ontological commitment](#).)

One particular argument for the ambiguity thesis runs as follows: 'Quine believes that there are no propositions' is true; and 'If Quine believes that there are no propositions, then there is a thing Quine believes' is a logical truth, and thus a truth. So 'There is a thing Quine believes' is true. But Quine is not ontologically committed to the existence of things believed (propositions) since he says that he does not think there are any. So in English 'there is' is not always an objectual quantifier. So, presumably, it is sometimes a substitutional one. Another argument observes that to explain why

(8) If there are exactly as many boys as girls, then if there are 5 boys, there are 5 girls

is logically true, one must interpret its antecedent as involving numerical quantification; but asserting its antecedent surely does not commit one to the existence of numbers (see [Gottlieb 1980](#)). The arguments depend on the neutrality thesis: a substitutional quantification ΣXFX does not commit one to there being an F if none of its instances do. Given the definition of ontological commitment, the neutrality thesis is trivially true, since any substitution instance of ΣXFX logically implies it.

Such arguments must be evaluated on a case-by-case basis. As for the first, one might simply note that existential generalization (the move from something whose syntactic form is ' a is F ' to 'Something is F ') is not valid without restriction. The second argument seems inconclusive, at least given that what is at issue is the commitments of one who embraces the range of quantitative inferences commonly held valid in English, and accepts arithmetic practice as well as standard use of numerical talk for measurement, counting, estimating, and so on.

Even if the evidence for the ambiguity thesis is at best equivocal, the project of attempting to see how much of a stretch of a discourse (about, say, numbers, or properties, or propositions) can be interpreted substitutionally is often of interest. Even if we are most plausibly understood to be committed to existence of a sort, we might welcome the news that we need not be. In evaluating the claim that a stretch of discourse can be interpreted substitutionally, one needs to consider several questions.

First, how satisfactory is the account provided of the truth of substitution instances? If 'Some numbers are prime' is substitutional, the truth of '7 is prime' is not explained by saying that 'is prime' is true of 7. We then need an account of what makes '7 is prime' true but '8 is prime' false.

Second, how much apparently true discourse does a substitutional account render problematic or false? It is not altogether clear how much conventional mathematics can be accepted when its quantifiers are read substitutionally. (For one thing, substitutional quantification seems to require limits on expressibility akin to those in predicative logic.) It is not clear how a substitutional interpretation of 'Katya said something' can explain its truth as a report of Katya's assertive utterance of a sentence untranslatable into English. (One could say that some set theory or talk of Katya's assertions is false but useful fiction.)

Third, to what extent do the terms of a substitutional account undercut its motivation? For nominalists, substitutional quantification is of interest because it may offer a way to re-interpret a theory (without requiring significant change in the way it is used) so that it can be adopted by the ontologically scrupulous. Such nominalists need to be careful about what they themselves (objectually) quantify over, in giving a substitutional account of a theory. For example, the natural understanding of the substitution class for ' Π ' for a language for arithmetic will identify it with an infinite set of expression 'types'; to give an adequate substitutional account of a language as strong as second-order arithmetic seems to require quantifying over sequences. But types and sequences are not objects acceptable to most nominalists (see [Parsons 1971](#) and [1982](#)).

Quine has claimed that substitutional quantification is irrelevant to questions of ontology, saying that 'the question of the ontological commitment of a theory does not properly arise' ([1969](#)) unless the theory is formulated in an objectual first-order language or translated into such. But surely this is not so. If quantifier-free expressions of English number theory do not in fact contain names of numbers, and English numerical quantification is substitutional, then English numerical discourse is free of commitment to numbers. In fact, if this is so, it is presumably untranslatable into objectual idiom, since such translation would involve wrongly ascribing such commitment.

3 Logic and semantics

There are many interesting issues in the logic and semantics of substitutional languages. The reader is directed to [Kripke](#), [Parsons](#) and [Grover](#) for a sampling. Here we confine ourselves to mention of two fundamental facts about substitutional languages.

First, $\Sigma X\phi X$ is true provided one of its substitution instances is. Thus, if one is to define truth inductively for a language in which substitutional quantifiers occur, with the definition assigning a truth-value to every sentence, the substitution class for the quantifiers cannot contain expressions in which those very quantifiers are used.

Kripke suggests that it is best to think of substitutional quantifiers as being introduced by extending a given (interpreted) language to a new language. The extension is achieved by introducing new variables for the quantifiers, defining a substitution class *C* thereof (consisting of some expressions of the original language) and taking the atomic sentences of the new language to be all the sentences of the old language along with results of replacing members of *C* in the base language with variables. (Call the last, following **Kripke**, ‘forms’. Strictly, forms are those results, of replacing members of *C* with variables, which always return a sentence of the base language when the new variables are replaced with members of *C*.) As **Kripke** (and, earlier, **Marcus**) observe, provided the quantifiers are introduced in this way, one can formulate ‘truth-definitions’ along the lines of (7) without paradox.

Second, in the case of first-order languages (without identity), choosing between objectual and substitutional quantifiers need not affect what we say about validity or satisfiability. (Again we rely on **Kripke**’s work.) Consider the first-order substitutional and first-order objectual predicate calculi, which differ only in their quantifiers and variables. Think of an interpretation of the substitutional calculus (SC) as being given by specifying an interpreted language *L* and (non-empty) substitution class *C* therein, assigning members of *C* to SC’s constants, forms (of appropriate number of argument places) of *L* to the predicate letters of SC. If validity for SC is truth in any such interpretation, and validity for the objectual predicate calculus is truth in any (objectual) interpretation, then a sentence of SC is valid if and only if its objectual analogue (the result of trading substitutional quantifiers and variables for objectual ones) is.

See also: [Logical and mathematical terms, glossary of](#); [Ontological commitment](#)

MARK RICHARD

References and further reading

- Gottlieb, D.** (1980) *Ontological Economy: Substitutional Quantification and Mathematics*, Oxford: Oxford University Press. (Expounds and defends a substitutional interpretation of number theory. Includes a useful discussion of ontological commitment.)
- Grover, D.** (1992) *A Prosentential Theory of Truth*, Princeton, NJ: Princeton University Press. (Uses substitutional quantification in characterizing truth for various languages.)
- Kaplan, D.** (1986) ‘Opacity’, in P. Schliipp (ed.) *The Philosophy of W.V. Quine*, La Salle, IL: Open Court. (An important discussion of quantification and opacity.)
- Kripke, S.A.** (1976) ‘Is There a Problem about Substitutional Quantification?’, in G. Evans and J. McDowell (eds) *Truth and Meaning*, Oxford: Oxford University Press. (Gives what has become the paradigmatic elementary account of substitutional quantification.)
- Marcus, R.B.** (1993) *Modalities: Philosophical Essays*, Oxford: Oxford University Press. (Collected essays. Numbers 1, 5 and 8 defend the substitutional understanding of English quantification as well as its ontological importance.)
- Parsons, C.** (1971) ‘A Plea for Substitutional Quantification’, *Journal of Philosophy* 68: 231-7. (Argues that substitutional quantification is not ontologically neutral.)
- Parsons, C.** (1982) ‘Substitutional Quantification and Mathematics’, *British Journal for the Philosophy of Science* 33: 409-21. (A review of **Gottlieb 1980**.)
- Quine, W.V.** (1951) ‘Reference and Modality’, in *From a Logical Point of View: Nine Logico-Philosophical Essays*, Cambridge, MA: Harvard University Press, 1980; repr. in L. Linsky (ed.) *Reference and Modality*, Oxford: Oxford University Press, 1971. (One source of Quine’s objections to objectual quantification into opaque positions.)
- Quine, W.V.** (1969) ‘Existence and Quantification’, in *Ontological Relativity and Other Essays*, New York: Columbia University Press, 1977. (Gives Quine’s objections to the relevance of substitutional quantification to ontology.)
- Quine, W.V.** (1973) *The Roots of Reference*, La Salle, IL: Open Court. (Contains a subtle discussion of substitutional quantification, mathematical discourse and the conceptual priority of substitutional quantification.)
- Richard, M.** (1995) ‘Propositional Quantification’, in J. Copeland (ed.) *Logic and Reality: Essays in Pure and Applied Logic*, Oxford: Oxford University Press. (Includes a critical discussion of substitutional accounts of propositional quantification.)

Quantum logic

The topic of quantum logic was introduced by Birkhoff and von Neumann (1936), who described the formal properties of a certain algebraic system associated with quantum theory. To avoid begging questions, it is convenient to use the term ‘logic’ broadly enough to cover any algebraic system with formal characteristics similar to the standard sentential calculus. In that sense it is uncontroversial that there is a logic of experimental questions (for example, ‘Is the particle in region R?’ or ‘Do the particles have opposite spins?’) associated with any physical system. Having introduced this logic for quantum theory, we may ask how it differs from the standard sentential calculus, the logic for the experimental questions in classical mechanics. The most notable difference is that the distributive laws fail, being replaced by a weaker law known as orthomodularity.

All this can be discussed without deciding whether quantum logic is a genuine logic, in the sense of a system of deduction. Putnam argued that quantum logic was indeed a genuine logic, because taking it as such solved various problems, notably that of reconciling the wave-like character of a beam of, say, electrons, as it passes through two slits, with the thesis that the electrons in the beam go through one or other of the two slits. If Putnam’s argument succeeds this would be a remarkable case of the empirical defeat of logical intuitions. Subsequent discussion, however, seems to have undermined his claim.

1 The logic of experimental questions

Idealizing physics somewhat, we can think of repeatable experiments as types of observation and the results as values of the corresponding observable. Given any set of possible results, R , for a type of observable, O , there is an experimental question, namely: ‘Will the observation of O produce a value in the set R ?’

If experiments are repeatable, then there must be some significance to the results of repeating them on systems in - idealizing again - the same state. For theories such as classical mechanics and quantum theory, therefore, we may take the statistics resulting from repetition to be manifestations of probabilities. In our idealized treatment, we then assign a precise numerical probability $u(a)$ to the ‘yes’ result for a given experimental question a when asked of a system in a state u . This connection between experimental questions and states imposes a logic on the experimental questions.

First we say that two experimental questions are ‘equivalent’ if they are assigned the same probability by every state. It is then convenient to talk as if equivalent experimental questions were identical. Next we define an entailment-like partial ordering, ‘ $\leq$ ’, on experimental questions: $a \leq b$ if, for all states u , $u(a) \leq u(b)$. The intuitive idea is that if a fails to ‘entail’ b then there is some state in which a ‘yes’ answer for a is more probable than a ‘yes’ for b . We define a maximum element 1 and a minimum element 0 corresponding to the trivial experimental questions which always give a ‘yes’ answer, or a ‘no’ answer, respectively. Moreover, there is a negation-like complement operator, where $a^\perp$, the complement of a , has a ‘yes’ answer precisely when a has a ‘no’ answer. Hence $(a^\perp)^\perp = a$ for any experimental question a . Also for any experimental question a and any state u , $u(a) + u(a^\perp) = 1$. From this it follows that the logic is ‘orthocomplemented’, that is, $a \leq b$ implies $b^\perp \leq a^\perp$. We may then define the ‘orthogonality’ relation $a \perp b$ (‘ a is orthogonal to b ’) to hold if $a \leq b^\perp$, or, equivalently, $b \leq a^\perp$. In standard logic two propositions are ‘inconsistent’ just in case one entails the negation of the other. Orthogonality is the analogue of inconsistency.

The ‘meet’ $a \wedge b$, if it exists, is the greatest (with respect to ‘ $\leq$ ’) experimental question less than or equal to both a and b . It is analogous to the conjunction $p \& q$ of two propositions, which is the meet with respect to the ordering of entailment. Likewise the ‘join’ $a \vee b$, if it exists, is the least experimental question greater than or equal to both a and b . It is analogous to the disjunction, p or q of two propositions, which is the join with respect to the ordering of entailment. Because the logic is orthocomplemented, De Morgan’s laws hold.

$$(a \wedge b)^\perp = a^\perp \vee b^\perp$$

$$(a \vee b)^\perp = a^\perp \wedge b^\perp$$

Thus far we have been exploring a priori the consequences of characterizing physics as having to do with idealized repeatable experiments on systems. The logic of experimental questions may be further developed by means of a

constraint which, although inspired by quantum theory and not merely the result of conceptual analysis, seems antecedently plausible. This is ‘postulate M’: given any pairwise orthogonal set of experimental questions $\{a_i\}$ there is an experimental question b such that, for any state u , $u(b) + u(a_1) + u(a_2) + \dots = 1$ (see Hughes 1989: 198). This is plausible because it amounts to there being an experimental question b which has the ‘yes’ answer just in case all the a_i have the ‘no’ answer. It may be inferred from postulate M by the Mackey-Maczynski theorem that the logic of experimental questions is ‘orthomodular’ (see Hughes 1989: 198), a weakened form of ‘distributive’. That is, every pairwise orthogonal countable subset has a join and the orthomodular property holds: for any experimental questions a and b such that $a \leq b$, $b = a \vee (b \wedge a^\perp)$.

We might expect the analogue of the material conditional, $(\text{not-}p) \text{ or } q$, to be $a^\perp \vee b$, but since the material conditional could also be characterized as $(\text{not-}p) \text{ or } (p \text{ and } q)$, another analogue is $a^\perp \vee (a \wedge b)$, which I abbreviate to $a \rightarrow b$. In fact it is the latter which is appropriate, because we require the analogue of *modus ponens*, namely $a \wedge (a \rightarrow b) \leq b$ (see Gibbins 1987: 139). The orthomodular property may then be stated succinctly: for any experimental questions a and b such that $a \leq b$, $b = a^\perp \rightarrow b$.

2 The nonclassical character of quantum logic

It is plausible, then, but not the result of pure conceptual analysis, that the logic of experimental questions forms an orthomodular partially ordered set.

For classical physics this logic is in addition a distributive lattice and so a Boolean algebra. (To say it is ‘distributive’ is to say that the familiar distributive laws hold.

$$a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c)$$

$$a \vee (b \wedge c) = (a \vee b) \wedge (a \vee c)$$

To say it is a ‘lattice’ is to say that every pair of experimental questions has a join and a meet.) The logic of experimental questions for quantum theory is not, however, a Boolean algebra. Instead it is usually taken to be isomorphic to an orthocomplemented partially ordered set which occurs naturally in an infinite-dimensional Hilbert space.

Without loss of generality we may think of a vector v in this Hilbert space as an infinite sequence of (complex) numbers $\langle v_1, v_2, v_3, \dots \rangle$. Given numbers s and t , we can form the linear combination $z = sv + tw$ of vectors v and w , with $z_i = sv_i + tw_i$ for all i . More generally, given any set e of vectors we can consider those vectors which are finite or infinite linear combinations of members of e . This is the (closed) ‘span’ of e , which we shall write as ‘ $[e]$ ’. (That it is closed with respect to the appropriate topology follows from the inclusion of infinite linear combinations.) A (closed) linear subspace is then any non-empty set a of vectors such that $[a] = a$.

These vectors are analogues of coordinate triples in three-dimensional Euclidean space. In this example, the linear subspaces, which are all closed, comprise (1) the coordinates of the origin, O , (2) lines through O , (3) planes through O , and (4) the whole of space. (1) and (4) correspond, respectively, to the elements 0 and 1 in the lattice, which is ordered by inclusion. So, for instance, if a is the set of coordinates for a line through O and b for a plane through O , then $a \leq b$ just in case the line is in the plane. The meet of two members of the lattice of closed linear subspaces is just their intersection. The join is not, however, the union, for the union is not in general a linear subspace (for example, the union of two lines is not a plane). Instead, the join is the *span* of the union (for instance, the set of coordinates of the plane containing the two lines).

Now some pairs of vectors in Hilbert space are orthogonal. This enables us to define the ‘orthogonal complement’ $e^\perp$ of a set of vectors e as the set of all vectors orthogonal to all vectors in e . This provides an alternative characterization of the span as $[e] = (e^\perp)^\perp$. Hence if a is a closed linear subspace, then $(a^\perp)^\perp = a$.

The lattice of closed linear subspaces of a Hilbert space is not a Boolean algebra, for the distributive laws fail to hold. The reason for this is easily seen in the case of the coordinate triples. Consider three different sets of coordinate triples a , b and c corresponding to three lines OA , OB and OC all in the one plane. Then $b \vee c$ corresponds to the plane they lie in, so $a = a \wedge (b \vee c)$. But $a \wedge b$ is the intersection of a and b and is therefore $\{ \langle 0, 0, 0 \rangle \}$. Hence $a \wedge b = 0$. Likewise $a \wedge c = 0$. So $(a \wedge b) \vee (a \wedge c) = 0 \vee 0 = 0$. The orthomodular law, a weakening of the distributive laws, does hold.

The claim that quantum logic is a lattice could be justified either by the convenience of the standard representation using closed linear subspaces of a Hilbert space, or on the basis that we should depart as little as possible from the classical case. It is not, however, easy to find any other reasons (see Hughes 1989: 200).

The nonclassical character of quantum logic is also exhibited by its valuations. Consider the following weakening of the requirements for a valuation of the standard sentential calculus: $\text{Val}(1) = 1$ and if $a \perp b$ then $\text{Val}(a \vee b) = \text{Val}(a) + \text{Val}(b)$, from which it follows that $\text{Val}(0) = 0$. Any (pure or mixed) quantum state may be thought of as such a valuation with $\text{Val}(a)$ being the probability assigned by the state to the experimental question a . It can be shown that such a valuation must take all the values between 0 and 1. (In particular, there is no two-valued or three-valued valuation. That there is no two-valued valuation is Kochen and Specker's theorem.) Following Kochen and Specker, consider a spin 1 system in which the spin component in any direction takes the values -1 , 0 and $+1$ (see Kochen and Specker 1967; Hughes 1989: 164-70). So for any direction p we have the experimental question o_p which has the 'yes' answer just in case the spin component in that direction is 0. For any three orthogonal p , q and r directions, the triad o_p , o_q and o_r are pairwise orthogonal experimental questions and their join is 1. Hence we would have $\text{Val}(o_p) + \text{Val}(o_q) + \text{Val}(o_r) = 1$. Gleason's theorem then tells us that $\text{Val}(o_p)$ is a quadratic function of the coordinates of p . Hence it is continuous and takes all values between the minimum, 0, and the maximum, 1.

3 An alternative approach to quantum logic

The presentation so far has been in the tradition of the paper by Birkhoff and von Neumann (1936) and the work of Mackey (1963). But there is an alternative approach to quantum logic due to Kochen and Specker (1965). The idea is to think of the logic as a family $\{B_i\}$ of overlapping Boolean algebras, each, intuitively, consisting of a family of experimental questions which can be answered without interfering with each other. We require the B_i to share the same maximum and minimum elements 0 and 1 (the two trivial questions) and to be such that the operations ' $\wedge_i$ ', ' $\vee_i$ ' and ' $\perp_i$ ' agree on the overlaps. The family $\{B_i\}$ is called a 'partial Boolean algebra' if a 'coherence condition' holds. Call members of the logic 'neighbours' if there is some B_i to which they all belong. Coherence then says that if a , b and c are pairwise neighbours then they are neighbours. The set of closed linear subspaces of a Hilbert space is a partial Boolean algebra in which the Boolean subalgebras consist of sets of subspaces which make angles of either 0° or 90° with each other. For the case of the lattice of coordinate triples, each Boolean subalgebra contains eight members and is specified by a choice of three perpendicular lines OP, OQ and OR through O. The subalgebra then consists of the coordinate sets for (1) the origin, O, (2) the lines OP, OQ and OR (3) the planes OQR, ORP and OPQ, and (4) the whole of three-dimensional space (see Hughes 1989: 192-4).

The natural way of defining a valuation for a partial Boolean algebra would be one which satisfied the usual truth-functional constraints on each Boolean subalgebra. But such a valuation would have to satisfy the constraint that $\text{Val}(1) = 1$ and that if $a \perp b$ then $\text{Val}(a \vee b) = \text{Val}(a) + \text{Val}(b)$. As we have seen, such valuations must take all values between 0 and 1.

4 The logical reformulation of quantum theory

The fact that one important difference between quantum and classical physics can be expressed as a purely algebraic difference in the corresponding logics has inspired the programme of formulating quantum theory in terms of its logic of experimental questions. There is no obstacle to this provided the logic is assumed to be that of the closed linear subspaces of a Hilbert space. Given that assumption, the first step is to classify the states, which are identified with the probability measures on the logic. Assuming postulate M (see §1 above) this amounts merely to taking the states to be valuations which assign values in the range 0 to 1. Gleason's theorem tells us that, for any Hilbert space of dimension greater than two, the states thus characterized are all the mixtures (in positive proportions adding up to 100 per cent) of the standard pure states of quantum theory, which correspond to vectors in the Hilbert space. (The state represented by vector v assigns to a closed linear subspace a the square of the cosine of the angle between v and a .)

The next step is to consider the dynamics by which states change over time. The results of Mackey and Kadison show that we may represent the vectors in the Hilbert space as suitable functions on ordinary (that is, three-dimensional Euclidean) space in such a way that the dynamics is given by the familiar Schrödinger equation (see Beltrametti and Cassinelli 1981: 252-4). The difficulty with the programme, however, is in providing simple

algebraic constraints which ensure that the logic of experimental questions is indeed represented by the set of closed subspaces of a Hilbert space (1981: ch. 21).

5 The status of quantum logic

What is the status of quantum logic? There seem to be three sorts of answer: the neoclassical, the quantum ontologists' and the quantum logicians'. The neoclassical answer is that quantum logic is more a mathematical curiosity than anything of great significance in understanding the physical world. Its non-Boolean character is the result of constraints on the questions it is physically possible for us human beings to answer, so that we might think of God as knowing the answers to two questions whose meet is the null question but which are not classically inconsistent (for example, the questions: 'Is the particle in bounded region R ?' and 'Does it have momentum in bounded set S ?'). Hence the meet, ' $\wedge$ ', would not be a genuine conjunction even though it has analogous properties. For that reason the partial Boolean algebra formulation might be judged more perspicuous because the ' $\wedge_i$ ' of a given Boolean subalgebra could be treated as a classical conjunction (see Garden 1984; Forrest 1988; Garola 1992).

The quantum ontologists' answer is that quantum logic tells us something important about the structure of the physical world, but that 'logic' is a misnomer, for it should not be interpreted as a theory of deductive reasoning. On this view the best way to describe the physical world is by specifying the 'logic' (better 'algebra') of experimental questions and specifying the states as probability measures on this logic (see §4). This is arguably implicit in the Copenhagen interpretation, and it would seem to be the position of van Fraassen (1991: ch. 5) and, in a more radical version, of Hughes (1989: ch. 10).

The quantum logicians' answer is that quantum logic holds because it is in fact the correct deductive logic for all thought, and we made a mistake in assuming classical logic, which we should abandon, not merely when considering quantum theory, but quite generally. This was Putnam's position and it has also been advocated by Stairs (see Putnam 1969; Stairs 1982).

How might we decide between these answers? While logic might well be empirically defeasible it does not seem to be initially derived from experience. Hence there is a burden of argument on the quantum logicians. Putnam (1969) took up this burden by arguing that abandoning the distributive law enables us to resolve various problems without abandoning a realist interpretation of quantum theory. If so, then indeed a case would have been made for the empirical defeat of classical logic, comparable to the empirical defeat of Euclidean geometry. Unfortunately, Putnam's claims are open to serious criticism. For example, Putnam noted that in the two-slit thought-experiment we might have expected the probability of a particle hitting the screen in a given region when both slits are open to be the mean of the two probabilities with one slit closed. But that result violates quantum theory. It seemed that we could, by rejecting distributivity, retain an interpretation in which particles really do pass through one slit or the other, without allowing action at a distance. The objection to this, made forcibly by Gibbins, is that even given quantum logic, realism about particles which pass through one rather than another slit, combined with the rejection of action at a distance, still gives us a result contrary to quantum theory, namely that the probability of a particle hitting the screen in a certain region should be at least as great as the mean of the probabilities with one slit closed. But in fact, for some regions that inequality does not hold (see Gibbins 1987: ch. 10).

If we reject the quantum logicians' answer then that of the quantum ontologists would be plausible enough, especially given the discovery of simple constraints on the logic which ensured it could be represented by the closed subspaces of a Hilbert space. Presumably, however, the choice between the neoclassical and the quantum ontologists' answers will depend on the overall advantages and disadvantages of their interpretations of quantum theory (see [Quantum mechanics, interpretation of](#)).

See also: [Logical and mathematical terms, glossary of](#)

PETER FORREST

References and further reading

- Bell, J.L.** (1986) 'A New Approach to Quantum Logic', *British Journal for Philosophy of Science*, 37: 83-99.
(Quantum logic is interpreted as a system of 'manifestations' of attributes in regions of a 'proximity space'.)
Beltrametti, E. and Cassinelli, G. (1981) *The Logic of Quantum Mechanics*, Reading, MA: Addison-Wesley. (A

- comprehensive guide to the mathematical foundations, especially useful for a discussion of the 'coordinatization problem', that is, the problem of representing quantum logic using Hilbert space.)
- Beltrametti, E. and Fraassen, B.C. van** (eds) (1981) *Current Issues in Quantum Logic*, New York: Plenum. (A fairly up-to-date, though technical, survey, with useful introductions to the topic by Mittelstadt and van Fraassen.)
- Birkhoff, G. and Neumann, J. von** (1936) 'The Logic of Quantum Mechanics', *Annals of Mathematics* 37: 823-43. (Of historical interest.)
- Dummett, M.E.** (1976) 'Is Logic Empirical?', in H.D. Lewis (ed.) *Contemporary British Philosophy*, London: Allen & Unwin, 45-68. (A critique of Putnam's claim that logic is empirical.)
- Forrest, P.** (1988) *Quantum Metaphysics*, Oxford: Blackwell. (Defence of the neoclassical position on quantum logic.)
- Fraassen, B.C. van** (1991) *Quantum Mechanics: An Empiricist View*, Oxford: Clarendon Press. (Careful and not too technical introduction to, among other things, quantum logic, with a case against the quantum logicians' position.)
- Garden, R.W.** (1984) *Modern Logic and Quantum Mechanics*, Bristol: Hilger. (Argues that the peculiarities of quantum theory are features of our own description rather than of the microscopic world.)
- Garola, C.** (1992) 'Truth Versus Testability in Quantum Logic', *Erkenntnis* 37: 197-221. (Shows how quantum logic can be obtained as the testable fragment of the standard sentential calculus.)
- Gibbins, P.** (1987) *Particles and Paradoxes: The Limits of Quantum Logic*, Cambridge: Cambridge University Press. (Lucid introduction to quantum logic, with an incisive critique of Putnam.)
- Hallam, N.** (1987) 'Logic and Indeterminacy', *Philosophical Papers* 16: 53-8. (The most recent paper in a debate between Harrison and Hallam, in which Harrison argues against the quantum logicians' position. See also their papers in *Analysis* of 1983-5.)
- Holdsworth, D.G. and Hooker, C.A.** (1983) 'A Critical Survey of Quantum Logic', in M.L. Dalla Chiara *et al.*, *Logic in the 20th Century: A Series of Papers on the Present State and Tendencies of Studies*, Aalen: Scientia, 127-246. (Useful survey of quantum logic with critical comments.)
- Hooker, C.** (ed.) (1975) *The Logico-Algebraic Approach to Quantum Mechanics*, Dordrecht: Reidel. (Anthology of papers, some hard to obtain in the original.)
- Hughes, R.** (1989) *The Structure and Interpretation of Quantum Mechanics*, Cambridge, MA: Harvard University Press. (An excellent introduction to quantum theory and its interpretation, including a not-too-technical discussion of quantum logic generally and the partial Boolean algebra approach in particular.)
- Kochen, S. and Specker, E.** (1965) 'Logical Structures Arising in Quantum Theory', in J. Addison, L. Henkin and A. Tarski (eds) *The Theory of Models*, Amsterdam: North Holland; repr. in C. Hooker (ed.) *The Logico-Algebraic Approach to Quantum Mechanics*, Dordrecht: Reidel, 1975, 263-76. (The paper introducing the partial Boolean algebra approach to quantum logic.)
- Kochen, S. and Specker, E.** (1967) 'The Problem of Hidden Variables in Quantum Mechanics', *Journal of Mathematics and Mechanics* 17: 59-87; repr. in C. Hooker (ed.) *The Logico-Algebraic Approach to Quantum Mechanics*, Dordrecht: Reidel, 1975, 293-328. (The original statement and proof of Kochen and Specker's theorem.)
- Ludwig, G.** (1983) *Foundations of Quantum Mechanics*, vol. 1, Berlin: Springer. (An operational approach to quantum theory.)
- Mackey, G.** (1963) *The Mathematical Foundations of Quantum Mechanics*, New York: Benjamin. (A historically important, but technical, treatment.)
- Mittelstadt, P.** (1986) 'Empiricism and Apriorism in the Foundations of Quantum Logic', *Synthese* 67: 497-525. (A discussion of the extent to which quantum logic has an empirical justification.)
- Putnam, H.** (1969) 'Is Logic Empirical?', in R. Cohen and M. Wartofsky (eds) *Boston Studies in the Philosophy of Science*, vol. 5, Dordrecht: Reidel; repr. in C. Hooker (ed.) *The Logico-Algebraic Approach to Quantum Mechanics*, Dordrecht: Reidel, 1975, 181-206. (Best-known philosophical case for the quantum logicians' position.)
- Putnam, H.** (1981) 'Quantum Mechanics and the Observer', *Erkenntnis* 16: 193-219. (This paper reflects some changes in Putnam's position.)
- Redhead, M.** (1987) *Incompleteness, Nonlocality, and Realism: A Prolegomenon to the Philosophy of Quantum Mechanics*, Oxford: Clarendon Press. (Careful and not-too-technical introduction to, among other things, quantum logic, with a discussion of Putnam's argument.)
- Stairs, A.** (1982) 'Quantum Logic and the Lurers' Rule', *Philosophy of Science* 49: 422-36. (Argument for the quantum logicians' position.)

Questions

Some theorists hold that a question is an interrogative sentence; others that a question is what is meant or expressed by an interrogative sentence. Most theorists hold that each question has two or more answers, and that the point of asking a question is to have the respondent reply with one of the answers. Most hold that each question has an assertive core or presupposition that is implied by each of the answers; if it is false, then no answer is true, so we say that the question commits the fallacy of many questions and we regard the negation of the presupposition as a corrective reply to the question (it corrects the question).

For example, consider the question 'Has Adam stopped sinning?' Its answers are 'Adam has stopped sinning' and 'Adam has not stopped sinning'. It presupposes 'Adam has sinned'; thus 'Adam has not sinned' is a corrective reply. The 'safe' way to ask this question is via the conditional 'If Adam has sinned, then has Adam stopped sinning?'

We can construct formal systems for asking whether and which questions in an effective way. Other types of question (for example, who and why) are still problematic. It can be proved that some questions are reducible to others, some questions raise others, and some systems for the logic of questions can never be complete in certain ways.

1 Concepts and theories

Questions have been discussed since Aristotle, although formal, systematic study did not begin until the 1950s. Theorists disagree on basic concepts, and there are at least three approaches to theorizing about questions. In one, the essence of a question is its set of answers, in another the essence is the questioner's intentions, and in another the essence is an objective intensional entity.

The 'reduction-to-answers' view holds that every question has direct answers, these direct answers are statements, and to know what counts as a direct answer to the question is to understand the question. A question presupposes any statement that is implied by every direct answer. A complete answer is a statement that implies a direct answer; a partial answer is a statement that is implied by a direct answer. The reduction-to-answers view is exemplified in most of the systems noted in §2 below.

The 'reduction-to-intentions' view holds that to know the questioner's intentions is to understand the question (see §4 below). The 'reduction-to-intensions' view holds that questions are intensional entities that are not relative to language and might or might not be expressible by sentences in a language (see §5 below and **Intensional entities**).

2 Effective systems

For many purposes (for example, conversation with strangers, courtroom procedure, automated information retrieval systems), it is desirable to have systems for question and answer that are effective in the following sense: for any expression of the given language, we can effectively tell whether it is an interrogative sentence and, if it is one, tell what question it expresses. In the reduction-to-answers view this implies that, given an interrogative sentence, we can effectively tell what counts as a direct answer to the question that is expressed.

Beginning with G. Stahl and T. Kubiński in the 1950s, logicians have constructed systems that provide for *whether* and *which* questions and that satisfy the effectiveness condition. The system of **Belnap and Steel (1976)** is exemplary. In it, *whether* and *which* questions are expressed by interrogatives of the form ?RS. Here the subject part S indicates a set of statements, which are the alternatives presented by the question. (For *whether* questions S is a finite list of statements; for *which* questions S is a formula indicating all the statements that result when names are substituted for the queriables (the free variables) in the formula.) Given S, the request part R indicates how many of the true alternatives are to be put into each direct answer and indicates whether the answer is to claim that it contains all of the true alternatives and (in the case of *which* questions) claim that the various names given in the answer denote distinct entities. The system provides for *who*, *when*, *where*, *why*, *what* questions - if these are construed as asking which person, which time, which place, which reason, which thing. It allows the questioner to call for a complete list or just some examples or an answer of the form at-least-*m*-but-at-most-*n*.

3 Some problems

Systems such as those noted above are good models for some question-asking situations but not for others. First consider ‘What is in the box?’ and ‘Some of John’s things are in the box’. If the latter is a complete answer, as some theorists say, then this *what* question cannot be construed as a *which things* question, and the system must be altered or extended to provide for it. Similar examples exist for *who* and other so-called ‘wh- questions’.

Why questions pose special problems. Consider ‘Why did this apple fall?’. We may construe this as ‘What is a good explanation of why this apple fell?’, but we will not know what counts as an answer until we know what counts as a good explanation. S. Bromberger (1992) has proposed that some types of *why* question can be adequately answered by citing (1) a general rule, (2) an abnormic law - that is, a law that specifies exceptions to the rule, and (3) some exception mentioned by the law. For example: ‘An apple remains on the tree unless it becomes ripe or a wind blows or the tree dies, and in this case the wind blew.’ It is reasonable to require that the abnormic law be empirical and true, but then (because we cannot always recognize truth) we cannot always recognize the admissible direct answers.

Effective systems have some intrinsic limitations. Suppose that the interrogatives of a system can be effectively recognized and hence listed by some algorithm and that, for each interrogative, its direct answers can be recognized and hence listed. Then it can be shown by Cantor’s diagonal method that the system is incomplete in the sense that there is a set of sentences that might be the answers to some question but cannot be the answers for any interrogative in the given list (Harrah 1984).

4 Epistemic analysis of questions

L. Åqvist (1965) proposed that questions are requests concerning epistemic states, expressible by imperatives such as ‘Let it be the case that I know that *A* or I know that *B*’ or ‘Let it be the case that, for some *x*, I know that *Px*’. In this approach there may be a variety of imperative operators, and reference made to different knowers, so that, for example, a teacher can put questions to a class in the form ‘Bring it about that I know that you know...’ J.

Hintikka (1983) has developed this approach further to elaborate its pragmatic dimension. Hintikka has extended his game-theoretic semantics to provide a semantics for epistemically construed questions and rules for conducting various types of rational question-and-answer dialogue (see *Semantics, game-theoretic*).

In this approach every question has a presupposition (in the first example above: ‘*A* or *B*’), a desideratum (‘I know that *A* or I know that *B*’) and direct answers (*A*, *B*). If an answer meets certain conditions of relevance and conclusiveness, guaranteeing that the desideratum is satisfied, then it is a conclusive answer. The conditions vary with the type of question. In the case of a *which* question, quantified variables should range over entities with which the questioner is acquainted, and the answer should specify a noun phrase only if the questioner knows what that phrase denotes. There is no guarantee that the respondent can effectively tell from the imperative sentence what the range of the variables is supposed to be, or what noun phrases will suffice, so, in this approach, the concept of conclusive answer in general is not effective.

5 Intensional analysis of questions

P. Tichý (1978) proposed that a question is a function defined on possible worlds. Propositions, individual concepts and properties are such functions, and they are the common types of question. For any given world the value of such a function is a truth-value, an individual or a set of individuals, respectively. In general the right answer to a question is the value of the function in the actual world.

This idea can be developed on its own; it can also be combined fruitfully with the ideas of Richard Montague. In Montague’s analysis of declaratives, meanings are higher-order intensional entities. These are assigned to linguistic expressions according to Frege’s principle that the meaning of a syntactical compound is the corresponding semantic compound of the meanings of the parts of the syntactical compound. Some linguists and logicians have developed analyses of this sort that treat both declaratives and interrogatives. In most, an interrogative expresses a question and denotes its true direct answers (see *Frege, G.*).

6 Erotetic logic

In a general sense erotetic logic is the logic of utterance and reply (Harrah 1984). In the strict sense it is the logic

of question and answer.

For most systems, especially those noted in §2 above, an important concept is that of containment. One question contains (or covers, or obviates) another if and only if every direct answer to the first implies some direct answer to the second, and two questions are equivalent if and only if each contains the other.

For analysis of the raising of questions, the implying of questions and arguing to questions, Wiśniewski (1995) has defined some useful concepts. For example: A set of declarative sentences evokes a question if and only if the set implies that the question has a true direct answer but does not imply the truth of any particular direct answer. Also, one question implies another if and only if each direct answer to the first implies that some direct answer to the second is true and for each direct answer to the second there is a proper subset (that is, a small selection) of direct answers to the first such that the given direct answer to the second implies that a true direct answer to the first is contained in that subset (so each true answer to the second helps us find a true answer to the first).

Another of Wiśniewski's concepts is that one question is reducible to a set of other questions (call it the reducing set) if and only if each direct answer to the given question implies that all the questions in the reducing set have true direct answers; and each collection of sentences that contains exactly one direct answer to each question in the reducing set implies some direct answer to the given question; and no question in the reducing set has more direct answers than the given question does. One result is that, if a question is 'safe' (that is, in every interpretation of the language the question has a true direct answer), then the question is reducible to some set of simple yes-no questions.

DAVID HARRAH

References and further reading

All the works cited contain some technical material. All contain useful bibliographies.

Åqvist, L. (1965) *A New Approach to the Logical Theory of Interrogatives*, Uppsala: Almqvist & Wiksell. (Referred to in §4. Outlines an approach based on commands about knowledge. Technical and difficult.)

Belnap, N.D., Jr and Steel, T.B., Jr (1976) *The Logic of Questions and Answers*, New Haven, CT: Yale University Press. (Referred to in §2. Presents many useful concepts.)

Bromberger, S. (1992) *On What We Know We Don't Know: Explanation, Theory, Linguistics, and How Questions Shape Them*, Chicago, IL: University of Chicago Press. (Referred to in §3. Good discussion of why questions.)

Groenendijk, J. and Stokhof, M. (1984) *Studies on the Semantics of Questions and the Pragmatics of Answers*, Amsterdam: Academisch Proefschrift. (Gives an intensional analysis with a pragmatic dimension.)

Harrah, D. (1984) 'The Logic of Questions', in D. Gabbay and F. Guenther (eds) *Handbook of Philosophical Logic, Vol. II: Extensions of Classical Logic*, Dordrecht: Reidel, 715-64. (Referred to in §§3, 6. Introduction, some history and a survey of various theories.)

Harrah, D. (1985) 'A Logic of Message and Reply', *Synthese* 65: 275-94. (§3 presents a general erotetic logic.)

Hintikka, J. (1983) 'New Foundations for a Theory of Questions and Answers', in Kiefer, F. (ed.) *Questions and Answers*, Dordrecht: Reidel, 1983, 159-90. (Referred to in §4. Outlines an approach based on a modification of Åqvist (1965).)

Hintikka, J. (ed.) (1988) 'Knowledge-Seeking by Questioning', *Synthese* 74: 1-262. (Papers by various authors, mainly on Hintikka's approach.)

Hintikka, J. (1992) 'The Interrogative Model of Inquiry as a General Theory of Argumentation', *Communication and Cognition* 25: 221-42. (Brief introduction to Hintikka-style systems.)

Kiefer, F. (ed.) (1983) *Questions and Answers*, Dordrecht: Reidel. (Papers presenting several approaches and discussing many aspects.)

Tichý, P. (1978) 'Questions, Answers, and Logic', *American Philosophical Quarterly* 15: 275-84. (Referred to in §5. Outlines an approach based on a theory of possible worlds.)

Wiśniewski, A. (1995) *The Posing of Questions: Logical Foundations of Erotetic Inferences*, Dordrecht: Kluwer. (Concerned mainly with raising and implying, as described in §6; technical and rigorous in some chapters but very clear throughout; has useful expositions, surveys and bibliography, and can serve as an introduction to the entire field.)

Randomness

The fundamental intuition underlying randomness is the absence of order or pattern. To cash out this intuition philosophers and scientists employ five approaches to randomness.

- (1) Randomness as the output of a chance process. Thus an event is random if it is the output of a chance process. Moreover, a sequence of events constitutes a random sample if all events in the sequence derive from a single chance process and no event in the sequence is influenced by the others.*
- (2) Randomness as mimicking chance. Statisticians frequently wish to obtain a random sample (in the sense of (1)) according to some specified probability distribution. Unfortunately, a chance process corresponding to this probability distribution may be hard to come by. In this case a statistician may employ a computer simulation to mimic the desired chance process (for example, a random number generator). Randomness qua mimicking chance is also known as pseudo-randomness.*
- (3) Randomness via mixing. Consider the following situation: particles are concentrated in some corner of a fluid; forces act on the fluid so that eventually the particles become thoroughly mixed throughout the fluid, reaching an equilibrium state. Here randomness is identified with the equilibrium state reached via mixing.*
- (4) Randomness as a measure of computational complexity. Computers are ideally suited for generating bit strings. The length of the shortest program that generates a given bit string, as well as the minimum time it takes for a program to generate the string, both assign measures of complexity to the strings. The higher the complexity, the more random the string.*
- (5) Randomness as pattern-breaking. Given a specified collection of patterns, an object is random if it breaks all the patterns in the collection. If, on the other hand, it fits at least one of the patterns in the collection, then it fails to be random.*

1 Chance

The most common conception of randomness identifies randomness with chance. Indeed, much of probability theory and statistics does not distinguish the two. Thus for a probabilist or statistician a *random event* and an *event due to chance* are typically the same thing. Moreover, processes giving rise to such events are referred to indiscriminately as *random*, *chance*, or *stochastic* processes.

Within statistics the adjective ‘random’ assumes a technical sense when it occurs in the phrase ‘random sample’. Given a chance process, one may wish to consider not just a single random event from this process, but rather an entire sequence of such events. Such a sequence is then said to constitute a random sample if: first, the same chance process is responsible for each event in the sequence; and, second, the occurrence of any event in the sequence is unaffected by the occurrence of other events in the sequence. If the first condition is satisfied one says that the events are ‘identically distributed’; if the second, that the events are ‘independent’.

Identifying randomness with chance now raises the obvious question, what is chance? While there exists a metaphysics of chance related to causation, determinism and free will, this entry approaches chance instrumentally in terms of those processes, like coin tossing and radioactive decay, for which our best understanding is irreducibly probabilistic (see [Chaos theory](#) for comparison, where the probabilities are artifacts of underlying deterministic systems).

2 Simulation

Scientific research consists increasingly of computer simulations that generate vast amounts of data. Presumably, if scientists had sufficient time and resources to examine nature directly, computer simulations that imitate nature would be unnecessary. Practical limitations, however, seem to render computer simulations indispensable to scientific research.

A dilemma now confronts the scientist. For many purposes the data a scientist wishes to obtain should properly be the output of a chance process characterized by some well-defined probability distribution. Practical limitations,

however, often prevent the scientist from actually sampling such a process and obtaining the desired data set (imagine a scientist who desires as data the sequence of heads and tails attained by flipping a coin one billion times - the scientist will expire before the billion flips can be accomplished). In this case, the scientist will want to simulate the chance process computationally.

The dilemma then is this. On the one hand computers are fully deterministic devices - specify an algorithm, and the behaviour of the machine is fixed. It follows that any probabilistic features of the data generated by a computer simulation are strictly eliminable. Yet, on the other hand, such data are to substitute for data generated by a genuine chance process, data which cannot be characterized except in probabilistic terms. As the output of a chance process, truly random data (in the sense of §1) are supposed to defy all but *post hoc* characterizations. As the output of an algorithm the (pseudo-) random data generated by a computer simulation are fully characterized in advance. How can the twain meet?

Strictly speaking they cannot. If randomness is identified with chance, then an event is random just in case it has the right sort of causal history and it was generated by a chance process. A computer is not a chance process. Ergo the data generated by a computer cannot be random. John von Neumann summed up the matter: ‘Anyone who considers arithmetical methods of producing random digits is, of course, in a state of sin’ (Knuth 1981: 1).

Nonetheless, the incongruity of using not merely deterministic systems, but systems whose entire behaviour can be precisely specified in advance has not dampened the proliferation of random number generators (RNGs). What then justifies using the data generated in a computer simulation in place of data generated by a chance process? In practice what happens is this. Given an RNG statisticians, as it were, set up a gauntlet of statistical tests that serve to vet it. The tests specify properties which the overwhelming majority of numerical sequences should have if they were generated by the chance process that the RNG is attempting to mimic. If the numerical sequences generated by the RNG do not have these properties, the RNG is rejected. Otherwise it is considered adequate.

Although in practice RNGs do a lot of useful work, there remains a theoretical problem in justifying RNGs in this way (which unfortunately is the only way RNGs can be justified): any RNG is only as good as the last statistical test that it passed. Indeed, history is strewn with RNGs that were for a time considered adequate, and then shown to be deficient. The problem is that we can never be sure that an RNG incorporates biases which the statistical tests we have thrown at it have simply failed to detect. This is bad. Practically speaking this means that the scientific literature may contain errors which we shall be unable to root out until appropriate statistical tests are found to detect the biases. For instance, cosmologists whose computer simulations of the early universe rely on RNGs may find their models overturned if the RNGs they employ are subsequently found to be badly biased.

3 Mixing

Take a fresh deck of playing cards and begin to riffle-shuffle them. How many riffle-shuffles are required before the deck is thoroughly mixed? Persi Diaconis has shown that seven riffle-shuffles are needed. What it means for the deck to be thoroughly mixed is that any configuration of the deck is as likely as any other. The deck starts in a specified configuration. A single shuffle mixes the deck, but not enough to break all connection with the previous configuration. Only after multiple shuffles does the configuration of the deck lose its connection with the starting configuration. At this point one says that the deck has attained a random state.

Shuffling a deck of cards is an example of a group action. Group actions provide one way of mixing things up, but not the only way. Imagine a gas concentrated in one corner of a box. The particles that make up the gas are in motion. Over time the gas will reach an equilibrium state, filling the entire box uniformly. This is an example of a dynamical system from statistical mechanics. The system starts out in a low entropy state, with the particles concentrated in one corner, and eventually reaches a maximal entropy state (an equilibrium state), with the particles evenly distributed throughout the box. The system is said to be random once it reaches the maximal entropy state (see *Thermodynamics*).

The preceding examples illustrate several features that are common to systems which attain randomness via mixing: (1) such a system starts out from a specified configuration that is highly ordered or constrained (for example, the opposite of what we would intuitively want to call random); (2) a mixing process (for example, a group action) acts on the system, over time continually transforming the configuration of the system; (3) eventually an equilibrium state is reached after which further mixing does not affect the equilibrium. When the

equilibrium state is reached, the system is said to be random.

It is worth noting that uniform probabilities frequently characterize the equilibrium states signalling randomness. What it means for a deck of playing cards to be thoroughly shuffled is that no configuration of the deck is more likely than any other. Shuffling has therefore randomized the deck only if each possible configuration of the deck is equiprobable. Similarly, a gas within a box has reached equilibrium if temperature throughout the box is uniform and the particles are evenly distributed. Uniform probabilities are therefore intimately connected with this understanding of randomness.

4 Complexity

Consider the following two sequences of a hundred coin tosses (heads = 1, tails = 0):

(R) 1100001101011000110111111
1010001100011011001110111
0001100100001011110111011
0011111010010100101011110

and

(N) 1111111111111111111111111
1111111111111111111111111
1111111111111111111111111
1111111111111111111111111

It seems clear that any theoretical account of randomness had better make (R) more random than (N). For instance, since (R) was obtained by actually flipping a coin whereas (N) was artificially contrived, according to the causal account of randomness sketched in §1, (R) would be random, but (N) nonrandom.

If, however, we prescind from our knowledge of the causal process responsible for these sequences, could we still distinguish these sequences in terms of randomness? We could, for instance, try to find a statistical test whose rejection region includes (N) and excludes (R), and thereby justify calling (N) nonrandom and (R) random. But for every such test it is possible to find a corresponding test whose rejection region includes (R) and excludes (N). Nor do probabilities help distinguish the sequences, since both (R) and (N), and indeed all such sequences of length 100, have the same small probability of occurring by chance, namely, 2^{-100} or approximately 1 in 10^{30} .

Starting in the 1960s a group of researchers that included Gregory Chaitin and Andrei Kolmogorov proposed a way around these difficulties (Earman 1986). Instead of characterizing randomness probabilistically, they took the very different tack of characterizing randomness computationally. What they said was that a string of 0s and 1s becomes increasingly random as the shortest computer program that generates the string becomes increasingly long. In the 1980s cryptographers proposed a variant of this characterization: a string of 0s and 1s becomes increasingly random as the most efficient computer program that generates the string requires increasingly long computation times. The first approach characterizes randomness in terms of space complexity (that is, the amount of memory the program occupies); the second, in terms of time complexity (that is, the computation time the program requires). The space complexity approach to randomness is referred to in the literature as ‘algorithmic information theory’.

It is now intuitively obvious why (R) is more random than (N). The shortest program that computes (N) has the form ‘repeat “1” 100 times’. On the other hand, (R) seems to have no shorter description than the string itself. (N) can be drastically compressed, (R) cannot. Thus from the point of view of algorithmic information theory (R) is more random than (N) (see **Information theory**).

Although complexity approaches to randomness represent a genuine advance in the theoretical study of randomness, there is a limitation to these approaches that is often lost in the initial enthusiasm: all complexity approaches to randomness are relativized to a given computational environment. What this means is that even though a sequence may be random when its generating program is running in PASCAL on a standard mainframe computer, with respect to another computational device it may be nonrandom, and vice versa. In fact, since mappings between finite sets are always computable (recursion theory on finite sets is trivial), any finite string will

be random with respect to certain programming environments, nonrandom with respect to others (see **Computability and information**; **Computability theory**).

5 Pattern-breaking

Having now surveyed four distinct approaches to randomness, one is tempted to ask whether a common thread runs through these approaches? There is a common thread, but one that at first sight will seem counterintuitive. In a dictionary definition of ‘randomness’, the term characterizes objects or events brought about without method, plan, purpose, forethought, pattern, principle, order or design. Random objects are supposed to be higgledy-piggledy, evincing no patterns.

But what does it mean for an object to evince no patterns? Consider a spy who eavesdrops on a communication channel in which encrypted messages are being relayed. If the spy has yet to break the cryptosystem, the encrypted messages traversing the communication channel will, as far as the spy is concerned, fail to display any patterns. Yet as soon as the cryptosystem is broken, all the patterns hidden by the cryptosystem become apparent.

The point is this: what determines the patterns that must be broken for an object to be random is not some objective feature of the world - randomness is not a natural kind. Rather, what is random depends on the patterns that are specified within a given context and that must then be broken for an object to be random. What is counterintuitive about this approach is that randomness becomes thoroughly parasitic on the patterns with respect to which it is defined. Randomness on this view does not make sense until a given collection of patterns is specified.

How then does this pattern-breaking approach to randomness relate to the four preceding approaches? For the computational complexity approach to randomness, the low complexity programs specify the patterns. For the mixing approach to randomness, far-from-equilibrium-states specify the patterns. For the simulation approach to randomness, statistical tests specify the patterns. The pattern-breaking approach to randomness also makes clear why chance is so often a dependable route to randomness: in many applications the patterns specified in advance identify a set of very small probability (for example, a full complement of statistical tests used to vet an RNG will typically designate as nonrandom only a tiny proportion of possible numerical sequences). Since small probability events are rare, chance will typically deliver objects or events that break all such patterns, that is, objects that are random in the pattern-breaking sense.

WILLIAM A. DEMBSKI

References and further reading

- Bayer, D. and Diaconis, P.** (1989) ‘Trailing the Dovetail Shuffle to its Lair’, Technical Report 329, Stanford, CA: Department of Statistics, Stanford University.(See §3.)
- Borel, E.** (1963) *Probability and Certainty*, trans. D. Scott, New York: Walker.(Treats the problem of small probabilities. See §5.)
- Dembski, W.A.** (1991) ‘Randomness by Design’, *Nous* 25 (1): 75-106.(See §5. An exposition of the pattern-breaking approach to randomness.)
- Diaconis, P.** (1988) *Group Representations in Probability and Statistics*, Hayward, CA: Institute of Mathematical Statistics.(The road to randomness via group actions. See §3.)
- Earman, J.** (1986) *A Primer on Determinism*, Dordrecht: Reidel.(Gives a general treatment of algorithmic information theory. See §4.)
- Garey, M.R. and Johnson, D.S.** (1979) *Computers and Intractability: A Guide to the Theory of NP-Completeness*, New York: Freeman.(An introduction to computational complexity in which computation time serves as the measure of complexity. See §4.)
- Hacking, I.** (1965) *Logic of Statistical Inference*, Cambridge: Cambridge University Press.(A good discussion of what statisticians mean by randomness in the sense of chance. See §1.)
- Knuth, D.E.** (1981) *Seminumerical Algorithms*, 2nd edn; in *The Art of Computer Programming*, vol. 2, Reading, MA: Addison-Wesley.(Includes Knuth’s classic treatment of random number generators. See §2.)
- Kranakis, E.** (1986) *Primality and Cryptography*, Stuttgart: Wiley-Teubner.(Treats random number generators as well as the time complexity approach to randomness. See §§2, 4.)
- Lambalgen, M. van** (1989) ‘Algorithmic Information Theory’, *Journal of Symbolic Logic* 53 (4): 1389-400. (Technical treatment of algorithmic information theory.)

- Lasota, A. and Mackey, M.C.** (1985) *Probabilistic Properties of Deterministic Systems*, Cambridge: Cambridge University Press.(The road to randomness via dynamical systems. See §3.)
- Mises, R. von** (1957) *Probability, Statistics, and Truth*, New York: Dover, 2nd edn.(Develops many seminal ideas about randomness; although much of what von Mises did has since had to be modified or corrected.)

Realism in the philosophy of mathematics

Mathematical realism is the view that the truths of mathematics are objective, which is to say that they are true independently of any human activities, beliefs or capacities. As the realist sees it, mathematics is the study of a body of necessary and unchanging facts, which it is the mathematician's task to discover, not to create. These form the subject matter of mathematical discourse: a mathematical statement is true just in case it accurately describes the mathematical facts.

An important form of mathematical realism is mathematical Platonism, the view that mathematics is about a collection of independently existing mathematical objects. Platonism is to be distinguished from the more general thesis of realism, since the objectivity of mathematical truth does not, at least not obviously, require the existence of distinctively mathematical objects.

Realism is in a fairly clear sense the 'natural' position in the philosophy of mathematics, since ordinary mathematical statements make no explicit reference to human activities, beliefs or capacities. Because of the naturalness of mathematical realism, reasons for embracing antirealism typically stem from perceived problems with realism. These potential problems concern our knowledge of mathematical truth, and the connection between mathematical truth and practice. The antirealist argues that the kinds of objective facts posited by the realist would be inaccessible to us, and would bear no clear relation to the procedures we have for determining the truth of mathematical statements. If this is right, then realism implies that mathematical knowledge is inexplicable. The challenge to the realist is to show that the objectivity of mathematical facts does not conflict with our knowledge of them, and to show in particular how our ordinary proof-procedures can inform us about these facts.

1 Platonism and intuitionism

Platonism, probably the most popular form of realism (and sometimes itself called 'realism'), is the view that mathematics is the study of a collection of distinctively mathematical objects which exist independently of us: numbers, functions, sets, lines and so on. Thus the Platonist takes mathematical discourse at face value: to say that there exists a prime number between seven and twelve is to make a straightforward existential claim; and to say that nine is greater than seven is to make a comparison between two objects. Mathematical objects, on the Platonist account, are abstract; they occupy no particular position in space or time. They are immutable and eternal, which explains the unchanging character of mathematical truth. They are related to one another in essential and also unchanging ways: the fact that nine is greater than seven, for instance, is no mere accident, but is part of the nature of the system of natural numbers.

On the Platonist account, reasoning about mathematical objects is just like reasoning about ordinary concrete objects, in that the objects, and collections of them, are simply 'out there' waiting to be investigated and described. There are of course many more mathematical than concrete objects, and the abstractness of the mathematical objects sets them apart, but they are none the less like concrete objects in their relation to our cognitive processes: in both cases the objects would exist, bearing the properties and relations they currently bear, whether or not we had ever given them any thought. And in both cases, it is reasonable to suppose that there are facts about the objects which we do not yet know, and even facts which we may be incapable of discovering.

At the other end of the realist/antirealist spectrum from Platonism is the position of 'intuitionism', according to which it makes no sense to speak of true-yet-undiscoverable mathematical facts. According to L.E.J. Brouwer, the twentieth century's paradigm intuitionist, our knowledge of mathematical objects is best explained by the view that these objects are actually *created*, not discovered, by us: mathematical objects are the products of human mental activity. As the intuitionist sees it, to say, for instance, that there is a natural number with such-and-such property is to say that we can mentally construct such a number. To say that every natural number has a given property is not to make a claim about some pre-existing infinite collection, but is rather to assert the existence of a particular kind of proof-procedure (see [Intuitionism](#); [Constructivism in mathematics](#)).

Reasoning about infinite collections in general is a much more cautious affair under the intuitionist's strictures than under the Platonist's, since the only true universal claims, according to the intuitionist, are those which come with effective proof-procedures. And because the only states of affairs relevant to the truth of mathematical

statements are those given by our own mental constructions, a statement is counted true by the intuitionist just in case it is provable. As a result, the kinds of mathematical reasoning accepted by the intuitionist differ from those accepted by the Platonist, and the intuitionist must reject some of the truths of classical mathematics.

Though the original motivation for intuitionism was a view about the creative powers of the human mind, a quite different motivation stems, in the work of Michael Dummett, from general views about the relationship between meaning, truth and verification (see [Dummett, M.A.E. §3](#)). As Dummett sees it, to grasp the meaning of a statement is to understand how it is used, and in particular to know what counts as a verification or falsification of it. Since mathematical statements are verified via mathematical proof, the view as applied to mathematics is that the meaning of a mathematical statement is given not in terms of facts which it purportedly represents, but rather in terms of the procedures which would count as a proof of it. Again, the upshot is that truth reduces to provability, and the idea of a realm of independent mathematical facts is rejected as incomprehensible.

The Platonist response to the intuitionist is that the revisionary picture recommended by the latter is under-motivated. Against Dummett, the argument is that we have a perfectly good conception of verification-transcendent truth-conditions, and hence of meaning which is given in terms of objective truth, and independently of provability. With respect to the original intuitionists' conception of mathematical objects, the Platonist's claim is that there is no reason to think of the objects of mathematics as dependent on us, and that the view of the mathematician's activity as one of creation, rather than discovery, does not do justice to this activity. An important complaint made by the Platonist in this regard is that the picture of mathematical objects as mental constructions fails to do justice to the communal nature of mathematical research, since it makes each mathematical object accessible only to its creator.

2 Platonism and epistemology

The strength of the Platonist position is that only the Platonist can hold that the ordinarily-accepted statements of classical mathematics are strictly and literally true. We all want to affirm, in some sense, that there is a prime number between three and six. But having said this, the non-Platonist must add when pressed that the claim is not to be taken quite literally, or is at least not to be taken to have objective truth-conditions. Only the Platonist can say what in weaker moments almost all of us would like to say, which is that there simply is a prime number between three and six.

But the literal reading of mathematical discourse does come at a price. Particularly problematic, at least on the face of it, is the explanation of mathematical knowledge. The mathematical objects posited by the Platonist are, unlike concrete objects, entirely divorced from human experience; we cannot perceive them, nor do they bring about any perceptible effects. It is therefore not easy to see how we might be in a position to know anything about them. A central challenge to the Platonist is to show that Platonism does not leave our mathematical knowledge a mystery.

There are two main lines of Platonist response to this epistemological challenge. The first is to claim that we bear a relationship to mathematical objects which is analogous to the perceptual relation we bear to concrete objects: as empirical data are to knowledge of concrete objects, so mathematical intuition is to knowledge of numbers, functions and sets. Such a position is held by Kurt Gödel ([1944](#)), who points out that even our knowledge of concrete objects is not direct, but is inferred from immediately evident sensations; so too, on this view, knowledge of mathematical objects is inferred from the immediate knowledge of simple mathematical truths. The primary challenge facing the mathematical-intuition theorist is to give a clear account of the nature of mathematical intuition, and of how, precisely, it relates us to the objects in question.

The second Platonist approach is to argue that knowledge about mathematical objects is simply knowledge of propositions, or perhaps of entire theories, in which those objects figure. The problem of explaining knowledge about the objects is then reduced to the problem of explaining the justification of those propositions or theories. Thus, for instance, Gottlob Frege ([1884](#)) holds that knowledge about numbers is simply knowledge of the propositions of arithmetic, which is explained in terms of more general capacities, and not in terms of individual contacts with the objects. An important question facing such an account is whether it is really Platonist; whether, that is, knowledge of propositions of the right form should count as, at bottom, knowledge about objects.

An explanation which attempts to reduce knowledge of mathematical objects to knowledge of entire theories is found in the work of W.V. Quine ([1960](#)) and Hilary Putnam ([1971](#)). On this account, we know about mathematical

objects in exactly the same way that we know about non-mathematical objects. Consider, for instance, the subatomic particles of modern physics. We know about these things, as Quine sees it, not in virtue of direct contact with them, but rather in virtue of the fact that our best theory includes the claim that these objects exist, and contains specific assertions about them. The evidence for the existence of these objects is just the empirical evidence we have for the truth of the theory as a whole. Similarly for ordinary concrete objects like tables and chairs, on this account: we posit these objects in order to make sense of the brute sensory data before us, and the success of the theory that posits them is the evidence we have that they exist. Given this picture of objectual knowledge, mathematical objects pose no special problem: the claim that there are such objects, and that they have the usual mathematical properties, is an indispensable part of our best theory of the world, since mathematics is an essential part of modern physics. Thus we have just as much reason to believe in the objects of mathematics as we do to believe in the existence of pieces of furniture. Whether such an account is successful will turn largely on the tenability of the holist account of knowledge in general.

An interesting response to such indispensability arguments is the ‘instrumentalist’ position, according to which the usefulness of mathematical statements in science is no evidence of their truth, and hence is no evidence of the existence of mathematical objects. As argued by Hartry Field (1980), the central claim is that the statements of mathematics offer a conservative extension of (say) physics, which is to say that, useful as mathematics is as a tool, it allows us to reach only those conclusions in physics which we would have been able to reach (albeit less efficiently) anyway. On this reading, the ordinary mathematical existence-assertions turn out to be useful but not true. The viability of this project turns on whether the conservativeness result can in fact be demonstrated and whether physics can in fact be reformulated in Field’s ‘nominalistic’ terms.

3 Anti-Platonist realism

A different response to the problems surrounding mathematical objects is to view the apparent object-directedness of mathematical discourse as just a mode of speaking, and not an indication that the truths of mathematics really are about objects. On this picture, mathematics is concerned rather with general features of collections, structures, potential measurements, or some such thing. Just as talk involving such phrases as ‘Emily’s sake’, or ‘the average family’ are easily seen to be rephrasable without the use of these terms, and to be free of commitment to any objects meeting the descriptions, so too statements purportedly about mathematical objects can be rephrased, it is argued, in such a way as to remove any *prima facie* commitment to such objects. As long as the rephrasings have objective truth-conditions which render them true or false independently of our ability to recognize whether or not they obtain, such an account is fully realist, though anti-Platonist.

One example of such a rephrasing is as follows. In recognition of the close connection between arithmetic and cardinality, one might hold that claims apparently about natural numbers can be cashed out as claims about collections and their sizes. Thus, for example, the sentence ‘Every natural number has a successor’ is taken to mean ‘For every possible finite collection, there is a possible collection with one more member’, where the phrase ‘one more’ is itself cashed out without reference to the number one. (The reference to possibility will be required in order for the claim to remain true even if the world happens to be finite.)

A more widely applicable reduction is the ‘structuralist’ account, on which mathematical statements are taken to be either about all collections of entities exhibiting the appropriate structure, or about those abstract structures themselves. Thus, the sentence ‘Every natural number has a successor’ can be taken to mean that for every (possible) collection whose members exhibit the pattern N (which we describe in terms of the successor-like relation S), for each member x of the collection there is a member y of that collection such that $S(x, y)$.

With even these simple examples, a number of important issues arise. The first is that the quantification over ‘possible collections’ may involve ontological commitments at least as strong as the commitment to numbers. Thus despite avoiding a commitment to numbers, one does not by such a strategy obviously obtain ontological parsimony, or avoid the difficulties associated with knowledge about abstract objects. Similarly, the usual ways of cashing out the necessary relations between structures or collections will involve quantification over functions; thus in the attempt to avoid one kind of mathematical object, the analysis is apparently committed immediately to another. On the other hand, such reductionist approaches have the benefit of construing arithmetical knowledge in a way which requires no epistemic relation to particular individuals (such as natural numbers), but only to whole classes of collections or patterns. This is an advantage if general properties of such collections or patterns are in

principle accessible to a priori reflection in a way in which distinguishing properties of individuals are not.

An important question about non-Platonist realism is whether it is coherent. Can one simultaneously maintain that the ordinarily accepted statements of mathematics are indeed true, but that there are no mathematical objects? The difficulty in this position is that it requires one to hold that statements such as ‘There are prime numbers between ten and twenty’ are true, while maintaining that there are no numbers. It is not enough to say that the quoted sentence can be rephrased in a way which apparently avoids mention of numbers. For if the rephrasing simply tells us what it means to say that there are prime numbers between ten and twenty, then its truth implies that there are such numbers. (See Alston (1958).)

It is tempting to try to avoid Platonism by saying that statements such as ‘There are prime numbers between ten and twenty’, and hence ‘There are numbers’ are true in some secondary sense - perhaps true ‘within mathematics’, but that they are not *really* true. But it is not clear whether this can be made sense of, since it is not clear what it is for a mathematical statement to be true in one sense and not in the other. In sum: a successful analysis of mathematical discourse which appears to eliminate reference to and quantification over particular mathematical objects does not by itself establish the claim that mathematical statements are free of commitment to mathematical objects. For all that has been said so far, such an analysis may well be regarded as an account of what it is to be committed to those objects. Whether there is more to Platonism than the acceptance of the objective truth of certain mathematical statements - and hence whether there is sense to be made of a non-Platonist realism - will turn, then, on the answers to some extra-mathematical questions concerning ontological commitment and the nature of objects in general.

See also: [Antirealism in the philosophy of mathematics](#); [Logical and mathematical terms, glossary of](#); [Objectivity](#); [Realism and antirealism](#)

PATRICIA A. BLANCHETTE

References and further reading

- Alston, W.P.** (1958) ‘Ontological Commitments’, *Philosophical Studies* 9: 8-17. (Argues that ontological commitment cannot be avoided by ‘translating away’ explicit reference to or quantification over purported entities.)
- Benacerraf, P.** (1973) ‘Mathematical Truth’, *Journal of Philosophy* 70: 661-80; repr. in P. Benacerraf and H. Putnam (eds) *Philosophy of Mathematics: Selected Readings*, Cambridge: Cambridge University Press, 2nd edn, 1983, 403-20. (A clear account of tensions between a Platonist account of mathematical content and acceptable theories of mathematical knowledge and reference.)
- Benacerraf, P. and Putnam, H.** (eds) (1964) *Philosophy of Mathematics: Selected Readings*, Cambridge: Cambridge University Press, 2nd edn, 1983. (A useful collection of central papers in the field.)
- Bernays, P.** (1935) Sur le platonisme dans les mathématiques, *L’Enseignement mathématique*, 1st series, 34: 52-69; trans. C.D. Parsons (1964), ‘On Platonism in Mathematics’, in P. Benacerraf and H. Putnam (eds) *Philosophy of Mathematics: Selected Readings*, Cambridge: Cambridge University Press, 2nd edn, 1983, 258-71. (Classic discussion of Platonism and its alternatives.)
- Brouwer, L.E.J.** (1912) *Intuitionisme et Formalisme*, Groningen: Noordhoff; trans. A. Dresden, ‘Intuitionism and Formalism’, *Bulletin of the American Mathematical Society* 20: 81-96, 1913; repr. in P. Benacerraf and H. Putnam (eds) *Philosophy of Mathematics: Selected Readings*, Cambridge: Cambridge University Press, 2nd edn, 1983, 77-89. (Gives an account of some of the central tenets of intuitionism.)
- Dedekind, R.** (1888) *Was Sind und was Sollen die Zahlen?*, Brunswick: Vieweg; trans. W.W. Beman, ‘The Nature and Meaning of Numbers’, in *Essays on the Theory of Numbers*, Chicago, IL: Open Court, 1901; repr. New York: Dover, 1963. (A classic of nineteenth-century foundations of mathematics. Gives an early non-eliminative but arguably structuralist account of the natural numbers.)
- Dummett, M.A.E.** (1967) ‘Platonism’, in *Truth and Other Enigmas*, Cambridge, MA: Harvard University Press, 1978, 202-14. (A criticism of the Platonist’s claim that our relation to mathematical objects is analogous to our perception of concrete objects.)
- Dummett, M.A.E.** (1973) ‘The Philosophical Basis of Intuitionist Logic’, in H.E. Rose and J.C. Shepherdson (eds) *Proceedings of the Logic Colloquium, Bristol, July 1973*, Amsterdam: North Holland, 1975, 5-40; repr. in *Truth and Other Enigmas*, Cambridge, MA: Harvard University Press, 1978, 215-47; and in P. Benacerraf and H. Putnam (eds) *Philosophy of Mathematics: Selected Readings*, Cambridge: Cambridge University Press, 2nd edn, 1983, 97-129. (Argues, from general principles about the theory of meaning, that classical logic should

be replaced by intuitionistic logic within mathematical reasoning.)

- Field, H.** (1980) *Science Without Numbers*, Oxford: Blackwell. (An argument that the usefulness of mathematics in physical science is no reason to hold that mathematical statements are true, as discussed in §2.)
- Frege, G.** (1884) *Die Grundlagen der Arithmetik: eine logisch-mathematische Untersuchung über den Begriff der Zahl*, Breslau: Koebner; trans. J.L. Austin, *The Foundations of Arithmetic: A Logico-Mathematical Enquiry into the Concept of Number*, Evanston, IL: Northwestern University Press, 2nd edn, 1980. (This classic statement of the logicist position is also an excellent introduction to many issues concerning mathematical realism. Defends the view that numbers are ‘self-subsistent objects’, as discussed in §2.)
- Gödel, K.** (1944) ‘Russell’s Mathematical Logic’, in P.A. Schilpp (ed.) *The Philosophy of Bertrand Russell*, Evanston, IL: Northwestern University Press, 125-53; repr. in Benacerraf, P. and Putnam, H. (eds) *Philosophy of Mathematics: Selected Readings*, Cambridge: Cambridge University Press, 2nd edn, 1983, 447-69. (Gödel’s view that the reasons for believing in mathematical objects are analogous to the reasons for believing in concrete objects, as discussed in §2.)
- Hart, W.D.** (ed.) (1996) *The Philosophy of Mathematics*, Oxford: Oxford University Press. (A collection of important recent papers.)
- Parsons, C.** (1990) ‘The Structuralist View of Mathematical Objects’, *Synthese* 84: 303-46; repr. in W.D. Hart (ed.) *The Philosophy of Mathematics*, Oxford: Oxford University Press, 1996. (A survey and critique of the structuralist treatment of mathematics.)
- Putnam, H.** (1971) *Philosophy of Logic*, New York: Harper. (Includes a clear statement of the ‘indispensability’ argument for the existence of mathematical objects, as discussed in §2.)
- Quine, W.V.** (1960) *Word and Object*, Cambridge, MA: MIT Press. (A classic. Includes the outline of Quine’s views on ontological commitment and theoretical posits, as discussed in §2.)
- Resnik, M.** (1981) ‘Mathematics as a Science of Patterns: Ontology and Reference’, *Noûs* 15 (4): 529-50. (An explanation and defence of a modern structuralist account of mathematics.)

Recursion-theoretic hierarchies

In mathematics, a hierarchy is a ‘bottom up’ system classifying entities of some particular sort, a system defined inductively, starting with a ‘basic’ class of such entities, with further (‘higher’) classes of such entities defined in terms of previously defined (‘lower’) classes. Such a classification reflects complexity in some respect, one entity being less complex than another if it appears ‘earlier’ (‘lower’) than that other. Many of the hierarchies studied by logicians construe complexity as complexity of definition, placing such hierarchies within the purview of model theory; but even such notions of complexity are closely tied to species of computational complexity, placing them also in the purview of recursion theory.

1 Hierarchies of formulas

Consider a formal language L with quantifiers $\forall$ and $\exists$ (see [Formal languages and systems](#)). Let B be a class of formulas of L such that by adding a $\forall$ - or $\exists$ -prefix to any member of B we get a non-member. We define a hierarchy of formulas of L that treats members of B as ‘basic’. Set $\prod_0^B = \sum_0^B = B$. Suppose that classes $\prod_n^B$ and $\sum_n^B$ have been defined. Let $\prod_{n+1}^B$ = the class of all formulas of L of the form $\forall v_1 \dots \forall v_m \varphi$ for $m \geq 1$, distinct variables $v_1, \dots, v_m$ and $\varphi \in \sum_n^B$; let $\sum_{n+1}^B$ = the class of all formulas of L of the form $\exists v_1 \dots \exists v_m \varphi$ for $m \geq 1$, distinct variables $v_1, \dots, v_m$, and $\varphi \in \prod_n^B$. This classifies all formulas of L that consist of a string of quantifier-prefixes attached to a member of B .

Motivation The complexity of a minimally complex formula (complexity understood as above) defining a given relation may be identified with the complexity of that relation. We can now ‘transfer’ this hierarchy to a hierarchy of the relations defined by such formulas over an appropriate structure. This was first done by S. Kleene (1943) with first-order formulas over the standard model of arithmetic to get the arithmetical hierarchy.

2 The arithmetical hierarchy

Let L_0 = the first-order language based on function-constants for successor, addition and multiplication, an individual-constant of 0, and a two-place predicate for $<$; variables range over ω . For $\varphi(x_0, \dots, x_m)$ a formula of L_0 with free-variables among those indicated, φ defines $\{ \langle a_0, \dots, a_m \rangle \in \omega^m : \varphi(a_0, \dots, a_m) \}$. A relation R is arithmetic iff R is defined by some formula of L_0 .

Let a formula φ of L_0 be bounded iff each occurrence of $\forall$ in φ is in a context of the form $\forall v$ (if $v < t$ then ...) and each occurrence of $\exists$ in φ is in a context of the form $\exists v$ ($v < t$ and ...), for t a term and v a variable; take B = the set of bounded formulae, $\prod_n^0 = \prod_n^B$, $\sum_n^0 = \sum_n^B$. For any m , $n \in \omega$, $m \geq 1$, and any $R \subseteq \omega^m$, we will also let: $R \in \sum_n^0$ iff R is defined by some $\sum_n^0$ formula of L_0 ; $R \in \prod_n^0$ iff R is defined by some $\prod_n^0$ formula; $R \in \Delta_n^0$ iff $R \in \sum_n^0 \cap \prod_n^0$. Context should make it clear whether $\sum_n^0$ and $\prod_n^0$ is a set of formulas or a set of relations. This is the arithmetical hierarchy of relations; since any defining formula has an equivalent in prenex form, it classifies all arithmetical relations. It looks like this (with rightward = up):

$$\begin{array}{ccccccc} & \subset \Sigma_1^0 \subset & \subset \Sigma_2^0 \subset & \subset \Sigma_3^0 \subset & & & \\ \Sigma_0^0 = \Pi_0^0 = \Delta_0^0 \subset \Delta_1^0 & & \Delta_2^0 & & \Delta_3^0 & & \Delta_4^0 \dots \\ & \subset \Pi_1^0 \subset & \subset \Pi_2^0 \subset & \subset \Pi_3^0 \subset & & & \end{array}$$

Any tuple of natural numbers can be effectively coded by a single natural number; so we lose nothing by restricting the arithmetical hierarchy to reals. The ‘arithmetical hierarchy theorem’ states that each class contains infinitely many ‘new’ reals (that is, for each new $n \in \omega$ $\sum_{n+1}^0 - \prod_{n+1}^0$, $\prod_{n+1}^0 - \sum_{n+1}^0$ and $\Delta_{n+1}^0 - (\sum_n^0 \cup \prod_n^0)$ are all non-empty). The definitional complexity of an arithmetic relation is indicated by the class in which it ‘first’ appears.

This hierarchy is related to classification by recursion-theoretic properties. Most significantly: Δ_1^0 = the class of recursive relations; $\sum_1^0$ = the class of r.e. relations; Δ_2^0 = the class of relations recursive in Post’s set K (i.e. those of T-degree $\leq 0'$). H. Putnam and Y. Ershov independently proved that Δ_2^0 is also the class of ‘trial-and-error’ relations: R is trial-and-error iff there is a recursive $m + 1$ - place function f such that given any

$\vec{a} = \langle a_0, \dots, a_m \rangle$ guesses forever as to whether $\vec{a} \in R$ ($f(\vec{a}, k)$ = the k th guess), and eventually f guesses correctly for ever after. In fact, they found and studied a fine-grained hierarchy within Δ_2^0 (see Putnam 1965).

These facts generalize ‘up’ the hierarchy. For a real X , set $X^{(n+1)}$ = the jump of $X^{(n)}$, with $X^{(0)} = X$. We then have: $X \in \Delta_{n+1}^0$ iff $X \leq_T \{\}^{(n)}$ for $\{\}$ = the empty set; furthermore, $X \in \Sigma_{n+1}^0$ iff X is r.e. relative to $\{\}^{(n)}$. (H. Rogers credits these basic results to Kleene and Post; see Rogers 1967.)

For $R \subseteq \omega^m$ and real Y : $R \leq_1 Y$ iff there is a $1-1$ recursive function $f: \omega^m \rightarrow \omega$ so that for any $\vec{a} \in \omega^m$: $\vec{a} \in R$ iff $f(\vec{a}) \in Y$. 1 -reducibility (that is, $\leq_1$) is the strongest of the recursion-theoretic reducibilities. For $\mathcal{C}$ a class of relations on ω : Y is $\mathcal{C}$ -complete iff $Y \in \mathcal{C}$ and for any $R \in \mathcal{C}$, $R \leq_\infty Y$. Fact: $\{\}^{(n)}$ is Σ_n^0 -complete and $\omega - \{\}^{(n)}$ is Π_n^0 -complete for each $n \leq 1$. Let $T_{\Sigma_n^0}$ and $T_{\Pi_n^0}$ be the sets of all and only the true Σ_n^0 and Π_n^0 sentences respectively. These sets have Σ_n^0 and Π_n^0 definitions respectively. (There is no paradox here: ‘I am not a true Σ_n^0 sentence’ requires negation of the truth predicate for Σ_n^0 , making it Π_n^0 rather than Σ_n^0 , and thus trivially true!) In fact, these two sets are also Σ_n^0 -complete and Π_n^0 -complete respectively, and thus (by a theorem of J. Myhill) are recursively isomorphic to $\{\}^{(n)}$ and to $\omega - \{\}^{(n)}$ respectively; this expresses the sense in which the jump is the recursion-theoretic analog of $\exists$ in arithmetic, and connects recursion-theory to truth-theory (see Tarski’s definition of truth; Semantic paradoxes and theories of truth).

Let us contrast the arithmetical hierarchy with classification of reals by T-degree. Each real has a T-degree; the arithmetical hierarchy only classifies arithmetic reals. In one respect T-degrees are finer than the ‘compartments’ of the arithmetical hierarchy (that is, $\Sigma_n^0 - \Delta_n^0$ and $\Pi_n^0 - \Delta_n^0$ for $n > 0$, $\Delta_{n+1}^0 - (\Sigma_n^0 \cup \Pi_n^0)$ for $n \geq 0$): for example, there are countably many T-degrees with members from $\Sigma_1^0 - \Delta_1^0$ (see Post’s problem in Turing reducibility and Turing degrees). On the other hand, for a real $X \in \Sigma_1^0$ with $\deg(X) = a$, $\omega - X \in a$; but if $X \notin \Delta_1^0$ then $\omega - X \notin \Sigma_1^0$; so except for the T-degree 0 no T-degree is a subset of Σ_1^0 ; similarly for the other compartments of the arithmetical hierarchy. So in another respect these classifications are orthogonal.

3 The analytical hierarchy

Expand L_0 to L_1 by introducing infinitely many predicate- (that is, second-order) variables, allowing $\forall$ and $\exists$ to bind them; without loss of generality let us require them all to be one-place. A relation on ω is analytical iff it is defined by a formula (containing no free predicate-variables), taking each predicate-variable to range over the class of all reals, with individual-variables ranging over ω . (‘Analytical’ alludes to quantification over reals, as in ‘Analysis’ in its mathematical sense.)

Now take B = the class of formula of L_1 containing no bound predicate-variables and no free individual-variables (though by coding tricks we could restrict B further). Set $\Sigma_0^1 = \Sigma_0^B$, $\Pi_0^1 = \Pi_0^B$. For any m , $n \in \omega$, $m \geq 1$, and any $R \subseteq \omega^m$, let: $R \in \Sigma_n^1$ iff R is defined by some Σ_n^1 formula of L_1 (with predicate variables ranging as above); $R \in \Pi_n^1$ iff R is defined by some Π_n^1 formula: $R \in \Delta_n^1$ iff $R \in \Sigma_n^1 \cap \Pi_n^1$; again, context disambiguates this double-usage. This is the analytical hierarchy. Using prenexing, it classifies all analytical relations. A picture of this hierarchy with regard to subsethood would look just like that of the arithmetic hierarchy, with superscripts changed from ‘0’ to ‘1’. Again, each class contains infinitely many new reals, there are Σ_n^1 -complete and Π_n^1 -complete reals, and among those are the truth-set for Σ_n^1 and Π_n^1 respectively.

4 Two more hierarchies of relation on ω

Set RA_0 = the set of arithmetic relations on ω . Suppose RA_α has been defined, for α an ordinal. A formula $\varphi(x_0, \dots, x_m)$ (all free variables first-order and among those shown) defines R over RA_α iff for any $\vec{a} \in \omega^m$: $\vec{a} \in R$ iff $\varphi(a_0, \dots, a_m)$, with predicate-variable ranging over RA_α ; note this change from §3. Let $RA_{\alpha+1}$ = the set of relations defined by such formulas over RA_α . For a limit ordinal λ : RA_λ = the union of all RA_α for $\alpha < \lambda$. This sequence is the ramified-analytical (hereafter RA) hierarchy: ‘ramified’ because the definitions used allow predicate-variables to range only over entities that were ‘previously’ defined. Note: if $\alpha \leq \beta$ then $RA_\alpha \subseteq RA_\beta$. This hierarchy is to reals what Gödel’s constructible hierarchy is to arbitrary sets (see Constructible universe); indeed work by R. Jensen shows that the latter can be construed as a recursion-theoretic

hierarchy (see [Devlin 1973](#)). Note one difference between these hierarchies: the RA hierarchy ‘peters out’ (at the countable ordinal $\beta_0 =$ least ordinal β such that $RA_{\beta+1} = RA_\beta$). In a natural sense the constructible relations on ω appearing the β_0 -stage of the constructible hierarchy extend the RA hierarchy; see [Odifreddi \(1989\)](#) for details.

The RA hierarchy can be further refined by breaking up each stage $RA_{\alpha+1} - RA_\alpha$ in terms of the quantifier-prefix of the defining formulae (in prenex-form, of course): $\sum_{\omega \cdot \alpha + n}^0$ = the set of relations defined over RA_α by $\sum_n^1$ formulas of L_1 ; now $\prod_{\omega \cdot \alpha + n}^0$ = the set of relations defined over RA_α by $\prod_n^1$ formulas; $\Delta_{\omega \cdot \alpha + n}^0 = \sum_{\omega \cdot \alpha + n}^0 \cap \prod_{\omega \cdot \alpha + n}^0$. For $n \geq 1$ we again have a hierarchy theorem: each such class contains infinitely many new reals. There are connections between the RA hierarchy, the jump operation, and truth in arithmetic.

An ordinal is recursive if it is order-isomorphic to some recursive well-ordering of natural numbers. There is a countable ordinal ω_1^{CK} = the least non-recursive ordinal (‘ CK ’ for Church and Kleene, who first studied it). $\omega_1^{CK} < \beta_0$, and if we cut off the RA hierarchy at ω_1^{CK} we get the hyperarithmetical (hereafter HA) hierarchy, introduced independently by M. Davis, A. Mostowski and S. Kleene (in [Davis 1950](#), [Mostowski 1951](#), [Kleene 1955](#)). For $\alpha = \omega_1^{CK}$, RA_α = the class of hyperarithmetical relations.

The so-called Souslin/Kleene theorem connects the HA hierarchy to the analytical hierarchy: Δ_1^1 = the class of the hyperarithmetical relations; so the HA hierarchy is a classification of the Δ_1^1 relations. Δ_1^1 can be thought of as the class of relations on ω that would be computable by a super-mind capable of completing an effectively specified ω -sequence of tasks, in which case Δ_1^1 is to Δ_1^0 as $\prod_1^1$ is to $\sum_1^0$; this analogy was one of the starting points for generalized recursion theory. The class of ramified-analytical sets (= RA_α for $\alpha = \beta_0$) is a proper subset of Δ_2^1 .

5 Hierarchies of relations between reals

So far we have only considered relations on ω defined by formulae of L_1 ; but a formula of L_1 whose only free-variables are one-place predicate-variables defines a relation between reals, that is, on $\text{Power}(\omega)$ = the class of reals. Such a relation is analytical iff it is defined in this way, allowing bound predicate-variables to range over $\text{Power}(\omega)$; it is arithmetical iff it is defined by such a formula with no bound predicate-variables (which formula may be classified like one of L_0). Thus we have an arithmetical and analytical hierarchy of such relations; for $i \in \{0, 1\}$, $R \subseteq \text{Power}(\omega)^m$: $R \in \sum_n^i$ iff R is so defined by a $\sum_n^i$ formula of L_1 ; $R \in \prod_n^i$ iff R is so defined by a $\prod_n^i$ formula.

Similarly, if we use ‘ramified’ definitions we obtain the HA hierarchy, and, going further, the RA hierarchy, of relations between reals. In the 1950s J. Addison (in his 1954 doctoral dissertation and in [Addison \(1959\)](#)) pointed out that the HA and analytical hierarchies of such relations were the ‘effective analogues’ of the so-called Borel and projective hierarchies, studied by French and Russian analysts (for example, E. Borel, H. Lebesgue, M. Suslin and N. Lusin) in the early twentieth century. Their work, usually labelled ‘descriptive set-theory’, was thought of as a part of topology; apparently the logicians (except for Mostowski) had worked in complete ignorance of it. Contemporary descriptive set-theory emerged from this cross-pollination as a hybrid of recursion theory and topology.

6 Other hierarchies

There is at least one interesting hierarchy within the class of recursive relations, the so-called Grzegorzcyk hierarchy. Even further ‘low-down’ (within the class of primitive-recursive relations) hierarchies have been proposed by computer-scientists, though here as this writing (1993) ignorance outstrips knowledge (see [Complexity, computational](#)). The study of such animals is usually called ‘subrecursion theory’.

All hierarchies discussed here classify relation on ω or on $\text{Power}(\omega)$. Practitioners of generalized recursion theory have generalized some of these hierarchies to classify relations on other classes of mathematical objects.

See also: [Logical and mathematical terms, glossary of](#)

HAROLD HODES

References and further reading

Addison, J. (1959) ‘Separation Principles in the Hierarchies of Classical and Effective Descriptive Set Theory’,

- Fundamenta Mathematica* 46: 123-135.(Apparently the first bridge between recursion theory and descriptive set theory.)
- Davis, M.** (1950) *On the Theory of Recursive Unsolvability*, Ph.D. thesis, Princeton University.(The first definition of the HA hierarchy.)
- Devlin, K.** (1973) 'Aspects of Constructibility', *Lecture Notes in Mathematics* no. 354, Berlin: Springer Verlag. (For the Jensen hierarchy, in effect Gödel's Constructible hierarchy of sets viewed recursion-theoretically.)
- Hinman, P.** (1978) *Recursion-Theoretic Hierarchies*, Berlin: Springer-Verlag.(A good text introducing a lot of this subject to the novice.)
- Kleene, S.** (1943) 'Recursive Predicates and Quantifiers', *Translations of the American Mathematical Society* 53: 41-73; repr. in M. Davis (ed.) *The Undecidable*, Raven Press, 1965, 255-287.(The first work on the arithmetical hierarchy.)
- Kleene, S.** (1955) 'Hierarchies of Number-theoretic Predicates', *Bulletin of the American Mathematical Society* 61: 193-213.(Kleene's first work on the HA hierarchy.)
- Mosschovakis, Y.** (1980) *Descriptive Set Theory*, Amsterdam: North-Holland Publishing Company.(An excellent text, connecting descriptive set theory and recursion theory.)
- Mostowski, A.** (1951) 'A Classification of Logical Systems', *Studia Philosophia* 4: 237- 274.(Another first paper on the HA hierarchy.)
- Odifreddi, P.** (1989) *Classical Recursion Theory*, Amsterdam: North-Holland Publishing. (An excellent introductory text on classical recursion theory as of 1989, with a lot on the arithmetical and analytical hierarchies, and some on the others.)
- Putnam, H.** (1965) 'Trial and Error Predicates and the Solution to a Problem of Mostowski', *Journal of Symbolic Logic* 30: 49-57.(A hierarchy within Δ_2^0 .)
- Rogers, H.** (1967) *The Theory of Recursive Functions and Effective Computability*, New York: McGraw Hill Inc. (The original textbook on recursion theory, a good complement to Odifreddi (1989), with excellent exercises.)

Reference

It is usual to think that referential relations hold between language and thoughts on one hand, and the world on the other. The most striking example of such a relation is the naming relation, which holds between the name 'Socrates' and the famous philosopher Socrates. Indeed, some philosophers in effect restrict the vague word 'reference' to the naming relation, or something similar. Others use 'reference' broadly (as it is used in this entry) to cover a range of semantically significant relations that hold between various sorts of terms and the world: between 'philosopher' and all philosophers, for example. Other words used for one or other of these relations include 'designation', 'denotation', 'signification', 'application' and 'satisfaction'.

Philosophers often are interested in reference because they take it to be the core of meaning. Thus, the fact that 'Socrates' refers to that famous philosopher is the core of the name's meaning and hence of its contribution to the meaning of any sentence - for example, 'Socrates is wise' - that contains the name. The name's referent contributes to the sentence's meaning by contributing to its truth-condition: 'Socrates is wise' is true if and only if the object referred to by 'Socrates' is wise.

The first question that arises about the reference of a term is: what does the term refer to? Sometimes the answer seems obvious - for example, 'Socrates' refers to the famous philosopher - although even the obvious answer has been denied on occasions. On other occasions, the answer is not obvious. Does 'wise' refer to the property wisdom, the set of wise things, or each and every wise thing? Clearly, answers to this should be influenced by one's ontology, or general view of what exists. Thus, a nominalist who thinks that properties do not really exist, and that talk of them is a mere manner of speaking, would not take 'wise' to refer to the property wisdom.

The central question about reference is: in virtue of what does a term have its reference? Answering this requires a theory that explains the term's relation to its referent. There has been a great surge of interest in theories of reference in this century.

What used to be the most popular theory about the reference of proper names arose from the views of Gottlob Frege and Bertrand Russell and became known as 'the description theory'. According to this theory, the meaning of a name is given by a definite description - an expression of the form 'the F' - that competent speakers associate with the name; thus, the meaning of 'Aristotle' might be given by 'the last great philosopher of antiquity'. So the answer to our central question would be that a name refers to a certain object because that object is picked out by the name's associated description.

Around 1970, several criticisms were made of the description theory by Saul Kripke and Keith Donnellan; in particular, they argued that a competent speaker usually does not have sufficient knowledge of the referent to associate a reference-determining description. Under their influence, many adopted 'the historical-causal theory' of names. According to this theory, a name refers to its bearer in virtue of standing in an appropriate causal relation to the bearer.

Description theories are popular also for words other than names. Similar responses were made to many of these theories in the 1970s. Thus, Kripke and Hilary Putnam rejected description theories of natural-kind terms like 'gold' and proposed historical-causal replacements.

Many other words (for example, adjectives, adverbs and verbs) seem to be referential. However we need not assume that all other words are. It seems preferable to see some words as syncategorematic, contributing structural elements rather than referents to the truth-conditions and meanings of sentences. Perhaps this is the right way to view words like 'not' and the quantifiers (like 'all', 'most' and 'few').

The referential roles of anaphoric (cross-referential) terms are intricate. These terms depend for their reference on other expressions in their verbal context. Sometimes they are what Peter Geach calls 'pronouns of laziness', going proxy for other expressions in the context; at other times they function like bound variables in logic. Geach's argument that every anaphoric term can be treated in one of these two ways was challenged by Gareth Evans.

Finally, there has been an interest in 'naturalizing' reference, explaining it in scientifically acceptable terms.

Attempted explanations have appealed to one or more of three causal relations between words and the world: historical, reliable and teleological.

1 Millian and description theories of proper names

The description theory of proper names stands in sharp contrast to the age-old and attractive theory that there is no more to a name's meaning than its role of designating something. Thus, John Stuart Mill claimed that 'proper names are not connotative: they denote the individuals who are called by them; but they do not indicate or imply any attributes as belonging to those individuals' ([1843] 1867: 20) (see [Mill, J.S. §2](#)). One problem for this theory is that it makes answering our central question seem hard: if a name does not 'imply any attributes' of its bearer, as the description theory claims it does, what determines which object is the bearer?

It was not primarily concern with this question, however, that led most philosophers to abandon the Millian theory and adopt the description theory. They did this largely as a result of the criticisms and counter-proposals of Frege (1893) and Russell (1911).

The most famous criticism of the Millian theory concerns identity statements. The ancient Greeks observed what they took to be a star rising in the evening and called it 'Hesperus', and what they took to be another star rising in the morning and called it 'Phosphorus'. In fact these 'two stars' were the planet Venus. Consider the following two statements:

- (1) Hesperus is Phosphorus.
- (2) Hesperus is Hesperus.

Sentence (1) is true. A comparison of (1) and (2) seems to indicate that they differ sharply in meaning. Various reasons have been adduced in favour of this view. Some philosophers have argued that whereas (1) is synthetic, (2) is analytic; others that whereas (1) is known empirically, (2) is known a priori; still others that whereas (1) is contingent, (2) is necessary. However, probably the most influential reason for thinking that the statements differ in meaning has been Frege's claim that they differ in 'cognitive value': (1) is highly informative, revealing an important astronomical discovery, whereas (2) is uninformative, a trivial piece of logical knowledge (see [Frege, G. §3](#)). In any case, it was generally agreed that the two statements do differ in meaning. If so, the only way to explain this seemed to be to attribute different meanings to 'Hesperus' and 'Phosphorus'. Yet, according to the Millian theory, they must have the same meaning as they both have the role of designating Venus. So the Millian theory must be wrong: a name's role of designating its bearer does not exhaust its meaning.

Another important criticism concerns existence statements. 'Vulcan does not exist' is true. Because it is true, 'Vulcan' does not designate anything. So, according to the Millian theory, 'Vulcan' should be meaningless. So, 'Vulcan does not exist' should be partly meaningless. Yet, it is perfectly meaningful. Indeed, if it were not, it could not be true. So, once again, the Millian theory must be wrong (see [Fiction, semantics of](#)).

The description theory provides neat solutions to what are problems for the Millian theory. According to the description theory, a name in effect abbreviates the definite description that competent speakers associate with the name: thus 'Aristotle' might abbreviate 'the last great philosopher of antiquity'. So it is easy to see why (1) and (2) differ in meaning: 'Hesperus' might be associated with the description 'the star that rises in the evening', and 'Phosphorus' with 'the star that rises in the morning'. And 'Vulcan does not exist' is fully meaningful because 'Vulcan' abbreviates the meaningful description 'the planet in orbit between Mercury and the Sun' (see [Proper names §§1, 5](#)).

More importantly for our purposes, the description theory provides an answer to our central question. A name designates a certain object because that object is denoted by the definite description associated with the name; 'Aristotle' designates Aristotle because 'the last great philosopher of antiquity' denotes him. Of course, this answer raises another question (see §6): in virtue of what does the description denote that object? Still, progress has clearly been made because we had that problem anyway.

Some obvious problems with this 'classical' description theory - including the problem that speakers differ in the descriptions that they associate with a name - led some philosophers, notably John Searle (1958) and Peter Strawson (1959), to modify the theory. A name is not tied tightly to one description but loosely to many. It can

designate its bearer despite the failure of some in its ‘cluster’ of associated descriptions to denote that object: it designates whatever object most of the descriptions in the cluster denote.

2 Three arguments against description theories of proper names

Description theories dominated for half a century until challenged by three arguments around 1970: the unwanted necessity and rigidity arguments, both due largely to Kripke (1980), and the argument from ignorance and error, due to Kripke and to Donnellan (1972).

Unwanted necessities were one of the obvious problems for the classical description theory. If ‘the last great philosopher of antiquity’ is synonymous with ‘Aristotle’, then ‘Aristotle is a philosopher’ should be necessarily true (provided Aristotle exists). Yet it is not: Aristotle might have died young, long before his philosophical fulfilment. The cluster theory avoided this version of the problem. The description ‘the last great philosopher of antiquity’ is just one among many in the cluster that expresses the meaning of ‘Aristotle’. Aristotle need not have any particular one of the many properties specified by the cluster. The cluster theory does require, however, that Aristotle have *most* of the properties specified by the cluster. Kripke points out how implausible this is. Aristotle might not have had any of the properties commonly associated with him: he might not have been a pupil of Plato, taught Alexander the Great, and so on. So ‘Aristotle’ cannot be synonymous with the cluster of associated descriptions (see Kripke, S.A. §2).

The rigidity argument deploys the notion of ‘rigid designation’. This is explained as follows: for a term *a* to be a rigid designator is for it to designate the same object in every possible world (in which it designates at all), or, less picturesquely, for it to be such that ‘*a* is *F*’ would truly characterize some non-actual situation if and only if the object that the term *actually* designates were *F* in that situation. Kripke argues that names are rigid designators whereas the descriptions alleged to be synonymous with them are not. So description theories are wrong. Consider:

- (3) Aristotle was fond of dogs.
- (4) The last great philosopher of antiquity was fond of dogs.

Suppose that Aristotle had indeed died young. Then Plato, not Aristotle, would have been the last great philosopher of antiquity. In those circumstances the truth of (4) would depend on whether Plato was fond of dogs. But the truth of (3) would still depend, just as it does depend in the actual world, on whether Aristotle was fond of dogs. The name ‘Aristotle’ designates Aristotle in a non-actual situation just as it does in the actual situation, whereas the description ‘the last great philosopher of antiquity’ designates whoever is the last great philosopher of antiquity in that situation, whether Aristotle or not. So the name is not synonymous with the description. Similarly, it is not synonymous with any other description, or cluster of descriptions, that is a candidate to give its meaning. (Note that in assessing rigidity we are evaluating the truth and reference of our expressions with the meanings that they actually have as a result of our usage, but we consider them as used to characterize hypothetical situations. Of course, any expression could have a different meaning as a result of different usage in a non-actual situation - language is ‘arbitrary’ - but that is beside the point) (see Proper names §§2-3).

Statements (3) and (4) are not modal statements (although we have been evaluating them as characterizations of non-actual situations). Other versions of the rigidity argument concern modal statements. For example, whereas ‘Hesperus is necessarily Hesperus’ is true, ‘Hesperus is necessarily the star that rises in the evening’ is not: had the solar system been differently arranged, Hesperus might not have been visible in the evening but it still would have been Hesperus. This sort of difference between descriptions and names in modal statements had been emphasized earlier by other philosophers, particularly Ruth Barcan Marcus (1961).

Some philosophers, notably Michael Dummett (1973), resisted Kripke’s two arguments by focusing on modal statements. These philosophers exploited the well-known ambiguities of scope in these statements to undermine the apparent difference between names and descriptions. Whatever the truth of this matter, the apparent difference in nonmodal statements remains.

These two arguments challenge the description theory as a theory of the meaning of a name, a meaning that determines the name’s reference. This is how the theory is naturally understood. However, as Kripke points out, the theory could be understood as simply a theory of reference: the reference of a name is fixed by a description, but the name is not synonymous with that description. This weaker theory is impervious to the two arguments. Of

course, the weak theory has a defect: because it is no longer a theory of meaning, it no longer solves the problems that troubled the Millian theory. Indeed, the relation between meaning and reference becomes a pressing issue on the basis of this theory.

There is another way of saving the description theory from the two arguments while avoiding this defect. Instead of weakening the original theory into a mere theory of reference, we revise it along the following lines: a name is synonymous with a 'rigidified' description. Our language already seems to have descriptions that contain 'rigidity operators'; for example, the italicized part of 'the person who, *in the actual world*, is the last great philosopher of antiquity' seems to make this description designate Aristotle in every possible world. If descriptions of this sort are indeed rigid, the revised theory claims that a name is synonymous with such a description. If such descriptions are not rigid, the revised theory can claim that the name itself supplies the rigidity operator and so would be synonymous with an ordinary nonrigid description governed by that operator.

All of these description theories - original, weak and revised - have the consequence that the users of a name associate with it a description that identifies its bearer. The third argument against description theories, the argument from ignorance and error, challenges this. So, if the argument is good, it counts against all description theories.

The argument shows that people who seem perfectly able to designate with a name are very often too ignorant to supply an identifying description. Thus some may fail with the name 'Cicero' because they associate with it only the description 'a famous Roman orator', which applies to many people. Others may fail because they associate 'the man who denounced Catiline' with 'Cicero' and are unable to supply an appropriate description for 'Catiline': the description that they associate with 'Catiline' is 'the man denounced by Cicero', which takes us in a circle and leaves both names without reference, according to the description theory.

The argument shows also that people often associate with a name a description that identifies something other than the name's bearer: they are simply wrong about the bearer. Thus some associate 'the inventor of the atomic bomb' with 'Einstein' and some associate 'the first person to realize that the world was round' with 'Columbus'. Almost everyone who has heard of Peano associates 'the discoverer of Peano's axioms' with 'Peano', but the axioms were actually discovered by Dedekind. Despite such errors, people succeed in designating Einstein, Columbus and Peano by their names.

The description theory can be improved by allowing people to 'borrow' their reference from others. So the description Martha associates with 'Einstein' might be 'the person Joe referred to yesterday as "Einstein"'. Provided Joe can supply an appropriate description - either one that describes Einstein directly or one that borrows reference from someone who can supply an appropriate description - Martha will succeed in designating Einstein. There is a danger of a circle, of course. Apart from that, there are problems of ignorance and error once more. Perhaps Martha cannot remember the reference lender; or she can remember the lender by his name, 'Joe', but cannot supply the identifying description that the theory requires; or the lender is identified but he cannot identify Einstein, perhaps identifying something else instead. The description theory still seems to place too great an epistemic burden on speakers.

An argument from ignorance and error can also be brought against another, more general, theory that some - for example, Dummett - have taken from Frege. This is the theory that to understand a name a person must be able to 'identify' its bearer. This ability is usually evidenced by providing a description, but it may be evidenced by 'recognizing' the bearer. The epistemic burden that this more general theory places on speakers still seems too great.

Various moves have been made to save the description theory in the face of these difficulties. Most popular, perhaps, have been theories that the reference of, for example, 'Einstein' is determined by a description along the lines of 'the person referred to by (called, named and so on) "Einstein"', for this description does identify Einstein and speakers surely associate it with the name. However, such theories still risk circularity.

3 General terms and mass terms

Just as there are description theories of names, so also there are description theories of general terms like 'tiger', 'hammer' and 'bachelor', and of mass terms like 'gold' and 'paper' (see **Mass terms**). Speakers of the language

associate various descriptions with a term. One of these descriptions, or most of a cluster of them, expresses the meaning of the term and determines what it applies to. If only one description does the job, the view is analogous to the classical description theory of names. If a cluster of descriptions does, the view is analogous to the cluster theory of names.

Kripke (1980) and Putnam (1975) argued that description theories are false of general and mass terms that apply to natural kinds. So they are false of 'tiger' and 'gold'. The arguments are like the three against description theories of names (see §2). First, the theories yield unwanted necessities. The description we associate with 'tiger' is along the lines of 'large carnivorous quadrupedal feline, tawny yellow in colour with blackish transverse stripes and white belly'. Yet it is not necessary that a tiger has four legs and is striped: a tiger might lose a leg, or in a different environment tigers might not be striped. Second, the term 'gold' is a rigid designator, applying to the same kind of stuff in every possible world. In contrast, an associated description like 'dense yellow metal' is nonrigid. Third, people who seem perfectly able to use a term are often too ignorant or misguided about the things to which it applies to supply an appropriate identifying description. Thus, some who use 'elm' and 'beech' cannot supply descriptions that distinguish elms from beeches; many who use 'gold' cannot distinguish gold from fool's gold; it was once common to associate 'fish' with 'whale'.

Putnam added a further argument built around the following fantasy. Imagine that somewhere in the galaxy there is a planet, Twin Earth. Twin Earth, as its name suggests, is very like Earth. In particular, each Earthling has a *doppelgänger* on Twin Earth who is a cell for cell duplicate of the Earthling. Twin Earth differs from Earth in one respect, however: the stuff that the Twin Earthians who appear to speak English call 'water', stuff that is superficially indistinguishable from what we call 'water', is not H₂O but a very different compound XYZ. So Oscar on Earth and Twin Oscar on Twin Earth are referring to different stuff by 'water'. Yet Oscar and Twin Oscar are *doppelgängers*, associating exactly the same descriptions with 'water' (which is more plausible if we place Oscar and Twin Oscar in 1750 before the chemical composition of water was known). So those associations are not sufficient to determine reference and the description theory is wrong. Indeed, nothing happening in the head is sufficient to determine reference. As Putnam put it, 'meanings just ain't in the head' (see [Content: wide and narrow](#); Putnam, H. §3).

We have considered criticisms of description theories of proper names and natural-kind words. Do these criticisms extend to description theories of other words? Putnam took the arguments to apply to almost all words, including 'pencil' and 'paediatrician'. Tyler Burge (1979) took a similar line, arguing that the meanings (contents) and references of a wide range of a person's words and accompanying thoughts are not 'individualistic' in that they are not determined simply by that person's intrinsic states. To a large extent they are determined by the person's social context. Burge's examples include 'arthritis', 'sofa', 'brisket', 'clavichord' and 'contract' (see [Methodological individualism](#)).

The Twin Earth fantasy brings out an important feature of description theories in general: even if a description theory gives the right answer to our central question for some word, its answer is incomplete. Thus, consider a description theory of 'tiger'. According to the theory, the reference of 'tiger' is determined by the reference of such words as 'carnivorous' and 'striped'. Suppose, contrary to the arguments above, that this were so. We then need to explain the reference of those words to complete the explanation of the reference of 'tiger'. Description theories might be offered again. But then the explanation will still be incomplete. At some point we must offer a theory of reference that does not make the reference of one word parasitic on that of other words. We need an 'ultimate' explanation of reference that relates some words directly to the world, if there is to be any reference at all.

4 Historical-causal theories

Kripke and Donnellan followed their criticism of description theories of names with an alternative view. This became known as the 'causal' or 'historical' theory, although Kripke and Donnellan regarded their view as more of a 'picture' than a theory (see [Proper names §4](#)).

The basic idea of this theory is that a name designates whatever is causally linked to it in an appropriate way, a way that does not require speakers to associate an identifying description of the bearer with the name. Reference is initially fixed at a dubbing, either by perception or description of the referent. The name is then passed on from

person to person in communicative exchanges. People succeed in designating an object with a name because underlying their uses of the name are causal chains stretching back to the dubbing of the object with the name. People borrow their reference from people earlier in the chain but borrowers do not have to remember lenders; it is enough that borrowers are, as a matter of historical fact, appropriately linked to their lenders in communication. So people can designate Cicero despite their ignorance of him, or designate Einstein despite their errors about him.

Similarly, Kripke and Putnam proposed an historical-causal theory of natural-kind words. Reference is initially fixed at a dubbing, either by description or perception of samples of the kind. The reference is then to all those objects, or all that stuff, having an internal structure of the same sort as the samples: for example, in the case of gold having the atomic number 79. People at a dubbing lend their reference to others, who can then lend it to still others. People who are ignorant about the kind can use the word to refer to the kind's members because underlying the use of the word are causal chains stretching back to a dubbing.

In thus removing the epistemic burden on speakers, historical-causal theories are a radical departure from the Frege-Russell tradition. That tradition assumes that those who understand a name must know about its meaning and reference, so that if its reference is determined in a certain way, they must know that it is. The historical-causal theory must reject the assumption that speakers have this privileged 'Cartesian' access to semantic facts: the reference of a name is determined by causal chains that are likely to be beyond the ken of the ordinary speaker. Once again, we must conclude with Putnam that 'meanings just ain't in the head'. This very feature of the theory has led many to reject it and to work hard to preserve the description theory (or the more general 'identification theory' favoured by Dummett; see §3). From the traditional Cartesian perspective, the causal theory's failure to impose an epistemic burden rules it out as a candidate to explain reference.

The historical-causal theory nicely captures the rigidity of names: the reference of a name is determined by its *actual* causal relations, something that cannot change when we consider other possible worlds. Less pleasingly, by rejecting any descriptive element to the meaning of a name, the theory may seem to leave no alternative but to resurrect the Millian view, identifying a name's meaning with its role of designating its bearer. Many philosophers, influenced by the 'direct reference' approach to indexicals (see §5), have taken this route, despite the problems for the Millian view (see §1). To avoid this it seems that we must explain a name's meaning in terms of the particular sort of causal chain that determines the name's reference.

The theory faces problems arising from various confusions and mistakes that can play a role in forming the causal network underlying a name. And it must explain how the reference of a name can change even though the historical fact of the dubbing cannot change. In developing the theory to deal with these problems, Michael Devitt (1981) has emphasized that a name is typically 'grounded' in its bearer in many perceptual confrontations after the initial dubbing: it is 'multiply' grounded in its bearer.

The historical-causal theory is returned to in §8.

5 Indexicals

To answer our central question for indexicals - terms like 'I', 'now', 'here', 'you', 'she', 'that' and 'this table' - we need to consider how the context of their utterance determines their reference.

Both Russell (1918) and Hans Reichenbach (1947) explain the reference of all indexicals (called 'egocentric particulars' by Russell and 'token-reflexive words' by Reichenbach) in terms of the reference of 'this'. Thus, Reichenbach claims that "'I" means the same as "the person who utters this token"; "now" means the same as "the time at which this token is uttered"; "this table" means the same as "the table pointed to by a gesture accompanying this token"' (1947: 284). This amounts to a description theory of reference for all indexicals except 'this token' in terms of the reference of 'this token'.

Any description theory of an indexical may face arguments of the usual three sorts: unwanted necessities, rigidity, and ignorance and error. The theory is particularly vulnerable to the argument from rigidity, as David Kaplan (1989) showed. (Kaplan prefers to talk of the closely related notion of 'direct reference' rather than 'rigidity'.) Thus, compare:

(5) This table is green

and Reichenbach's interpretation of it:

(6) The table pointed to by a gesture accompanying this token is green.

Suppose that the table referred to is in fact green, so that (5) and (6) both assert true propositions. Consider now a situation in which that table was still green but the furniture had been moved around so that a different table, a brown one, would be the subject of the gesture. Would those propositions asserted by (5) and (6) still be true? (Note that this question concerns the propositions actually asserted by (5) and (6), not the propositions that would have been asserted by the sentences in that non-actual situation.) Kaplan argues that whereas what (6) asserts would be false, what (5) asserts would still be true. The indexical 'this table' is rigid, referring to the same table in each possible world, whereas the description 'the table pointed to by a gesture accompanying this token' is nonrigid, referring to whatever table fits that description in the possible world. So the demonstrative is not synonymous with the description.

Apart from this, a general theoretical consideration counts against a description theory of indexicals. We have noted the essential incompleteness of description theories (see §3): even if a description theory is right for some word, the theory's explanation must rest ultimately on the reference of some other words which must be explained nondescriptively. Indexicals seem to be the most plausible candidates for nondescriptive explanation, more so even than proper names or natural-kind words: indexicals seem to be the place where language stands in its most direct relationship to the world.

In seeking a nondescriptive theory, it helps to follow Kaplan in dividing indexicals into two groups: 'pure indexicals' like 'I', 'here' and 'now', and 'demonstratives' like 'she', 'that' and 'this table'. The nondescriptive explanations of pure indexicals are fairly simple: 'I' designates the speaker of the utterance, 'here' the place of the utterance, 'now' the time of the utterance, and so on. (These explanations may seem to be description theories once again, but they are crucially different. For example, the last explanation is not that 'now' designates the time of the utterance because it is synonymous with an associated description 'the time of this utterance', but rather, because it is governed by the rule that it designates that time.) Demonstratives are more difficult to explain.

There are three basic ideas for a nondescriptive explanation of demonstratives. According to the first, a demonstrative designates the object demonstrated by the speaker. One problem with this idea is that a demonstration is often so vague that it alone would not distinguish one object from many others in the environment. A more serious problem is that demonstratives are not always accompanied by a demonstration. Thus, where only one table is prominent in the environment, the speaker may use 'this table' without a demonstration. And reference is often to an object that is not around to be demonstrated: for example, 'That drunk at the party last night was offensive'.

According to the second idea for a nondescriptive explanation, a demonstrative designates the object that the speaker intends to refer to. Even if this is so, it does not take us far because it raises the question: in virtue of what does the speaker intend to refer to that object? This is very similar to the original problem.

According to the third idea - urged, for example, by Edmund Husserl (1900-1) - a demonstrative designates the object in which it is based perceptually (compare the perceptually based grounding of a name according to the historical-causal theory). So 'this table' designates a certain table in virtue of the fact that it was perception of that table that led to the utterance; similarly 'that drunk at the party' designates the person that caused the remark (see [Demonstratives and indexicals](#)).

6 Descriptions

Definite descriptions have the form, 'the *F*', and indefinite descriptions the form, 'a/an *F*'. In his theory of descriptions, Russell (1905) claimed that 'the *F* is *G*' is equivalent to 'there is something that is alone in being an *F* and it is *G*'; and 'an *F* is *G*' is equivalent to 'there is something that is an *F* and it is *G*'. So the descriptions are to be understood in terms of the general term, '*F*' and the existential quantifier, 'there is something'.

Under the influence particularly of Donnellan (1966), many now think that a description is 'ambiguous', having not only this 'attributive' meaning captured by Russell but also a 'referential' meaning like that of a name or demonstrative.

It has been generally agreed that descriptions have a referential use as well as an attributive use. Used attributively, ‘the *F*’ conveys a thought about whatever is alone in being *F*; ‘an *F*’ conveys a thought about some *F* or other. Used referentially, each description conveys a thought about a particular *F* that the speaker has in mind, a thought about a certain *F*. Thus, consider:

(7) The murderer of Smith is insane

used in the following two contexts. (a) We come upon Smith foully murdered. We have no idea who is responsible but the brutal manner of the killing leads us to utter (7). Its description is used attributively. (b) We observe Jones on trial for Smith’s murder. The oddity of his behaviour leads us to utter (7). Its description is used referentially. Next, consider:

(8) A man in a red cap stole Anne’s computer

used in the following two contexts. (c) Anne’s computer is discovered missing in the morning. We find signs that the burglar made a hasty escape dropping a red cap in the alley. This leads us to utter (8). Its description is used attributively. (d) After discovering that Anne’s computer is missing we remember noticing a man in a red cap behaving suspiciously earlier in the day. We utter (8) to report our suspicions to the boss. Its description is used referentially.

Despite agreement that there are these two uses, there is no agreement that descriptions are ambiguous. Appealing to ideas prominent in the work of H.P. Grice (1989), many have defended Russell. They argue that a speaker can use a description referentially, thus making the object in mind the ‘speaker referent’, even though that object is not the semantic referent. Whether a speaker has an object in mind or not, the truth-conditions of the sentence are as specified by Russell. The referential use is pragmatically, but not semantically, different from the attributive use (see *Pragmatics* §§3-4, 8). Thus, in context (b), although Jones is the speaker referent, the truth of (7) will depend on the sanity of whoever murdered Smith, whether Jones or not. And in context (d), although the man behaving suspiciously is the speaker referent, the truth of (8) will depend on whether some man or other in a red cap stole the computer.

Against this, many have found reasons for thinking that in contexts like (b) and (d) the speaker referent is also the semantic referent and hence that descriptions are semantically ambiguous after all. Some of these reasons - for example, those arising from failures to describe correctly the object in mind and from the behaviour of descriptions in opaque contexts - have not stood up well. Others seem more promising. First, we are not just able to use descriptions referentially, it seems that we regularly do so. This regularity suggests that there is a convention of so using descriptions. If there is, then it is hard to see why the convention is not semantic. Second, in their referential uses descriptions seem to have roles just like demonstratives: ‘the *F*’ and ‘an *F*’ function like ‘that *F*’, and are similarly based on perception of a particular object. To try to treat these demonstratives like Russellian descriptions would be to give a description theory of them, and we have already noted problems for this (see §5). Third, definite (but not indefinite) descriptions seem to have the same range of anaphoric roles as a pronoun like ‘she’ (see §7). We might then expect them also to share the pronoun’s role as a demonstrative which, as has already been pointed out, seems not to be Russellian. Fourth, consider the utterance, ‘The book is on the table’. In the right circumstances, this will seem true and yet, according to the Russellian view, it must be false: since the world is full of books and tables the two definite descriptions fail to describe unique objects. The obvious modification to save the Russellian view is to treat these ‘incomplete’ descriptions as elliptical. But this modification has problems. A speaker may have many ways to complete the description and there may be no basis for saying that any one is the correct way. Alternatively, trying to complete the description may lead to the familiar problems of ignorance and error.

Another argument against the ambiguity thesis appeals to rigidity. If referential uses of descriptions were semantically significant then, it is claimed, they should be rigid like names and demonstratives. Yet they do not seem to be. Consider the use of ‘Smith’s murderer’ in context (b) above. ‘Smith’s murderer is insane’ does not seem to be true in a world where Smith is alive and well, even if Jones is insane. However, this argument has a problem: the referential use of the demonstrative ‘that murderer’ would equally seem to fail this rigidity test in these circumstances (hence suggesting a need to revise Kaplan’s claims about demonstratives; see §5). Yet a demonstrative surely has a semantically significant referential use. If so, then a description may have one too,

despite not being rigid (see [Descriptions](#)).

7 Other terms

Many terms that we have not discussed - like adjectives, adverbs and verbs - are naturally taken to refer. It is certainly no easier to explain reference for these terms than for the terms discussed, but we may hope that doing so will not pose sharply different problems (see [Adverbs](#); [Predication](#)).

We must also consider sentential operators like 'and' and 'not', and quantifiers like 'every pen', 'some stones', 'most dogs' and 'few bachelors'. Perhaps these should be seen as largely syncategorematic (see [Logical constants](#); [Quantifiers](#)). If we are prepared to accept the existence of certain abstract entities, however, we can take these expressions as referential also. Thus we can take 'and' as denoting a 'truth function' *conjunction* which is such that the sentence '*p* and *q*' is true if and only if '*p*' is true and '*q*' is true. The quantifiers involve a 'determiner' and a general term, and can be taken as applying to sets. Thus 'most dogs' involves the determiner 'most' and the general term 'dogs' and can be taken as applying to any set that contains more than half the dogs; and the sentence 'most dogs bark' is true if and only if there is such a set and 'bark' applies to all its members.

In virtue of what do these expressions have these referents? The most promising answer for the sentential operators has two stages. We start by describing the 'conceptual role' of the operator in deductive, inductive and practical inferences. For a token to denote 'conjunction' it must have the appropriate conceptual role. But in virtue of what should we assign to a token with that role the denotation 'conjunction' rather than, say, 'disjunction'? Because, under that assignment, deductive inferences are truth-preserving, inductive inferences are reliable, and so on. A similar line is presumably part of the answer for the quantifiers: 'most dogs' applies to any set containing most dogs (rather than, say, to any set containing a few cats) partly because of the reference of 'dogs' and partly because of the conceptual role of the determiner 'most' and the reliability of inferences. A worrying feature of both these answers is that they seem to make widespread irrationality impossible (see [Semantics](#), [conceptual role](#)).

Finally, we must consider anaphoric terms. Pronouns, and even definite descriptions, often depend for their reference on other expressions in their verbal context. Thus 'one' in 'John owns a car and Alice owns one too' is (using Geach's term) a 'pronoun of laziness', going proxy for the noun phrase 'a car' in the preceding conjunct. And consider 'he' in 'John is happiest when he is alone' and in 'Every man knows a woman that he admires'. In the former sentence 'he' is naturally seen as coreferential with 'John', in the latter as 'bound by' the quantifier 'every man' and so functioning like a bound variable in logic.

Geach (1962) has argued that all anaphoric pronouns are either pronouns of laziness or bound by quantifier antecedents. Against this Evans (1985) has argued that some pronouns with quantifier antecedents are unbound. He calls these 'E-type'. (a) Consider:

(9) Few congressmen admire the president, and they are very junior.

If 'they' were bound by 'few congressmen', (9) should mean that few congressmen both admire the president and are very junior. But it does not. The first clause of (9) entails that few congressmen admire the president; the second that all of those who do admire him are very junior. (b) If 'they' were bound in (9), then we should be able to substitute any quantifier for 'few congressmen' and still make sense. But 'No congressmen admire the president, and they are very junior' does not make sense. (c) Last, consider:

(10) If many men come to the ball, Mary will dance with them.

The quantifier 'many men' could bind 'them' only if (10) meant 'Many men are such that if they come to the ball Mary will dance with them', which is not the natural reading. As a result of these considerations and others - particularly pronouns in one person's sentence that are anaphoric on quantifiers in another person's - it is generally agreed that Evans has identified a distinct type of pronoun.

However, Evans' view of this type has been challenged. He thinks that the reference of such a pronoun is determined in a Russellian way by a definite description that can be derived from its quantified antecedent; thus, the reference of 'they' in (9) is determined by 'the few congressmen who admire the president', and that of 'them' in (10) by 'the many men who come to the ball'. Because these are definite descriptions, for (9) to be true *all* the Kennedy admirers must be junior, and for (10) to be true Mary must dance with *all* the men who come. This

consequence does not seem to generalize. ‘Some congressmen admire the president, and they are very junior’ seems to be compatible with some other congressmen admiring him and not being very junior. The problem is more acute in singular cases: ‘Socrates owned a dog and it bit him’ seems to be compatible with Socrates owning another dog which did not bite him. Finally, there are the formidably difficult ‘donkey sentences’: ‘Every man that owns a donkey beats it’ and ‘If John owns a donkey, he beats it’. On one reading, these sentences concern not simply the unique donkey of each donkey owner but *all* the owner’s donkeys (see [Anaphora](#)).

8 Naturalizing reference

From a naturalistic perspective, reference must ultimately be explained in scientifically acceptable terms. Attempted explanations have appealed to one or more of three causal relations between words and the world: historical, reliable and teleological.

Historical-causal theory of reference. Kripke, Donnellan and Putnam (see §4) did not claim to be naturalizing reference, but their theories - together with the role that perception of an object may play in determining the reference of demonstratives and referential descriptions (see §§5-6) - suggest the idea that reference might be explained naturalistically in historical-causal terms: a token refers to the object that played the appropriate role in causing it. But this idea, developed by Devitt, faces the *qua*-problem. In virtue of what is ‘Aristotle’, say, perceptually grounded in a ‘whole object’ and not a time-slice or undetached part of the object, each of which is equally present and causally efficacious? The problem is more pressing for natural-kind words. ‘Horse’ is grounded in a few horses. But those objects are not only horses, they are mammals, vertebrates and so on; they are members of very many natural kinds. Indeed, any horse is a member of indefinitely many *non*-natural kinds: it may be a pet, an investment and so on. In virtue of what is ‘horse’ grounded in an object *qua* horse, rather than *qua* mammal, pet, or whatever? So in virtue of what does it refer, as a result of such groundings, to all and only horses rather than all and only mammals, pets, or whatever?

Reliabilist theory of reference. Under the influence particularly of Fred Dretske (1981) and Jerry Fodor (1990), ‘reliabilist’, ‘informational’, or ‘indicator’ theories have been popular. The basic idea is that a token refers to objects of a certain sort because tokens of that type are reliably correlated with the presence of those objects; the tokens are ‘caused by’ those objects. The token ‘carries the information’ that a certain situation holds in much the same way that tree rings carry information about the age of a tree. There is a problem. How can the theory allow for error? Occasionally we see a muddy zebra and wrongly think ‘horse’. So, some zebras are among the things that would cause tokens of ‘horse’. What ‘horse’ is reliably correlated with is really the presence of horses, muddy zebras, the odd cow in bad light and so forth. So according to reliabilism, it should refer to horses, muddy zebras, the odd cow and so on (with the result that it was not wrong to think ‘horse’ after all). The problem is that many things that a token of a certain type does not refer to, including some denizens of Twin Earth, would cause a token of that type (see [Semantics, informational](#)).

Teleological theory of reference. Most fully developed by Ruth Millikan (1984), teleological theories explain the reference of a token in terms of its function, where that function is explained causally along Darwinian lines: a token’s function is what tokens of that type do that explains why they exist. This theory deals neatly with the problem of error because something - for example, sperm - can have a function which it does not reliably perform. An immediate consequence of the theory is that a token of a type that has not evolved will lack a referent. So the ‘thoughts’ and ‘utterances’ of an exact replica of Russell created by some cosmic accident would have no reference. This strikes many as implausible but is accepted by the theory’s proponents. To complete the theory it must be shown that tokens - even a belief like ‘computers make writing easier’ which could not plausibly be taken as innate - have a function in the required biological sense and that this function does indeed relate the token to its referent. Millikan has attempted this formidable task (see [Semantics, teleological](#)).

9 Further issues

Terms in opaque or intensional contexts cannot be seen as having their usual referential roles. For, in these contexts, particularly those of propositional attitude ascriptions (see [Propositional attitude statements](#)), the replacement of a term by a coreferential term may not preserve truth.

There are a range of what might be called ‘negative’ views of reference. (a) Some philosophers have a

'deflationary' view according to which there is nothing more to referential notions than is captured by all instances of a schema like '*e* designates *a*', where what is substituted for '*a*' 'translates' what is named by the term substituted for '*e*'; "'Socrates" designates Socrates' is a typical instance. This view accompanies a similarly deflationary view of truth (see [Truth, deflationary theories of](#)). (b) W.V. Quine (1960) argues that even once the translation of a sentence has been fixed the reference of any part of the sentence is inscrutable; thus there is no fact of the matter whether an alien's 'Gavagai' in response to an environment of rabbits refers to rabbits, undetached rabbit parts, time-slices of rabbits, and so on (see [Radical translation and radical interpretation §§1-4](#)). Related to this, Donald Davidson (1984) takes an instrumentalist attitude to reference, denying both the need for, and the possibility of, a theory of reference. Putnam (1983) gives a model-theoretic argument that reference is indeterminate because any theory has unintended models. (c) Kripke (1982) presents an argument (which he finds in Ludwig Wittgenstein's discussion of rule-following) that the meanings and references of terms are not determinate (see [Meaning and rule-following](#)). (d) Less sweepingly, Hartry Field (1973) has argued that in some cases there is no determinate matter of fact whether a term refers to one thing or another and we should see it as 'partially referring' to both: for example, 'mass' as used by Newtonians does not determinately refer to either proper mass or relativistic mass but partially refers to both. (e) Finally, those in the 'structuralist' tradition reject reference, and hence its role in meaning, altogether. They apparently think that the only possible theory of reference is one according to which a word resembles what it refers to. But this theory is refuted by the fact that language is arbitrary, by the fact that anything could be used to mean anything. Reference is thus left as simply 'God-given', which is unacceptable (see [Structuralism in linguistics](#)).

Finally, views on reference can bear on realist notions about the external world (see [Realism and antirealism](#)). Putnam, for example, draws antirealist conclusions from his model-theoretic argument. And consider the consequences of a holistic description theory for scientific terms. When, in time, we come to replace one scientific theory with another, it is natural to think that part of the reason we do so is that the theory does not accurately describe reality. Combine this thought with the holistic view that the reference of each term in the theory is determined by its associations with all other terms in the theory, and we get the consequence that all terms in the theory fail to refer. So, it was a mistake to believe in the entities apparently referred to by that theory. Worse, it is probably a mistake to believe in the entities of our present theory, for that theory will surely be replaced in time too. So we should not be scientific realists. Indeed, these considerations lead Thomas S. Kuhn (1970) and others to constructivism, a radically relativistic antirealism: rather than saying that the replaced theory does not describe reality, they say that it describes *its* reality, a reality that only exists relative to that theory (see [Constructivism](#)). Each theory has its own reality and no sense can be made of scientific entities existing 'absolutely'. This line of thought can be resisted by rejecting the holistic description theory of reference in favour of a localist theory, perhaps one explaining reference in terms of causal relations to reality (see [Holism: mental and semantic](#)).

See also: [Semantics](#)

MICHAEL DEVITT

References and further reading

- Burge, T.** (1979) 'Individualism and the Mental', in P.A. French, T.E. Uehling, Jr. and H.K. Wettstein (eds) *Midwest Studies in Philosophy*, vol. 10, *Studies in the Philosophy of Mind*, Minneapolis, MN: University of Minnesota Press. (Argues that reference for many terms is determined by the social context; referred to in §3.)
- Chastain, C.** (1975) 'Reference and Context', in K. Gunderson (ed.) *Minnesota Studies in the Philosophy of Science*, vol. 7, *Language, Mind and Knowledge*, Minneapolis, MN: University of Minnesota Press. (A long paper on singular reference and anaphora. A source for the view that indefinite descriptions have two uses.)
- Davidson, D.** (1984) *Inquiries into Truth and Interpretation*, Oxford: Oxford University Press, 215-25. (Argues against the need for, and possibility of, theories of reference; referred to in §9.)
- Devitt, M.** (1981) *Designation*, New York: Columbia University Press. (Develops a causal-historical theory of proper names, demonstratives and referential descriptions along naturalistic lines; referred to in §§4 and 8.)
- Devitt, M. and Sterelny, K.** (1998) *Language and Reality: An Introduction to the Philosophy of Language*, Oxford: Blackwell, 2nd edn. (Provides an introductory discussion of reference and of its significance for issues such as realism. Also includes an annotated guide to further reading.)
- Donnellan, K.S.** (1966) 'Reference and Definite Descriptions', *Philosophical Review* 75: 281-304. (Influential work proposing a distinction between referential and attributive descriptions and criticizing Russell and Strawson; referred to in §6.)

- Donnellan, K.S.** (1972) 'Proper Names and Identifying Descriptions', in D. Davidson and G. Harman (eds) *The Semantics of Natural Language*, Dordrecht: Reidel. (Criticizes the description theory of proper names and proposes a historical theory; referred to in §§2 and 4.)
- Dretske, F.** (1981) *Knowledge and the Flow of Information*, Cambridge, MA: MIT Press. (Most influential proposal of a reliabilist theory of reference; referred to in §8.)
- Dummett, M.A.E.** (1973) 'Appendix: Note on an Attempted Refutation of Frege', in *Frege: Philosophy of Language*, London: Duckworth. (A defence of a Fregean view of proper names from Kripke's criticisms; referred to in §2.)
- Evans, G.** (1982) *The Varieties of Reference*, ed. J. McDowell, Oxford: Clarendon Press. (Defends 'Russell's Principle' that to think about an object a person must know which object is in question.)
- Evans, G.** (1985) *Collected Papers*, Oxford: Clarendon Press. (Chapters 4, 5 and 8 argue for the view that some anaphoric pronouns are 'E-type' rather than bound; referred to in §7.)
- Field, H.** (1973) 'Theory Change and the Indeterminacy of Reference', *Journal of Philosophy* 70: 462-81. (Argues that the reference of scientific terms is sometimes indeterminate and that they should be seen as 'partially referring'; referred to in §9.)
- Fodor, J.A.** (1990) *A Theory of Content and Other Essays*, Cambridge, MA: MIT Press. (Chapters 3 and 4 propose a reliabilist theory of reference appealing to 'asymmetric counterfactuals'; referred to in §8.)
- Frege, G.** (1893) 'On Sense and Reference', in P. Geach and M. Black (eds) *Translations from the Philosophical Writings of Gottlob Frege*, Oxford: Blackwell, 1952. (Classic work raising the problem of identity statements for a Millian theory of proper names and proposing the theory that names have 'senses'; referred to in §1.)
- Geach, P.** (1962) *Reference and Generality*, Ithaca, NY: Cornell University Press. (Argues that all anaphoric pronouns are either like bound variables or are 'pronouns of laziness'; referred to in §7.)
- Grice, H.P.** (1989) *Studies in the Way of Words*, Cambridge, MA: Harvard University Press. (Proposes the distinction between speaker meaning and semantic meaning that has been influential in discussions of definite descriptions; referred to in §6.)
- Husserl, E.** (1900-1) *Logical Investigations*, trans. J.N. Findley, London: Routledge & Kegan Paul, 1970, part 1, §26; part 6, §§1-5. (Proposes a theory of demonstratives based on perception; referred to in §5.)
- Kaplan, D.** (1989) 'Demonstratives: An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives and Other Indexicals', in J. Almog, J. Perry and H. Wettstein (eds) *Themes from Kaplan*, New York: Oxford University Press. (Influential work arguing that the reference of a demonstrative is 'direct' not via a Fregean sense; referred to in §5.)
- Kripke, S.A.** (1980) *Naming and Necessity*, Cambridge, MA: Harvard University Press. (Very influential work criticizing the description theory of proper names and natural-kind terms and proposing a causal 'picture'; first presented in lectures in the late 1960s; referred to in §§2-4.)
- Kripke, S.A.** (1982) *Wittgenstein on Rules and Private Language: An Elementary Exposition*, Cambridge, MA: Harvard University Press. (Argues from a consideration of rule-following that the meanings and references of terms are not determinate; referred to in §9.)
- Kuhn, T.S.** (1970) *The Structure of Scientific Revolutions*, Chicago, IL: Chicago University Press, 2nd edn. (In this brilliant work on scientific change, a description theory of reference leads to antirealist constructivism; referred to in §9.)
- Marcus, R.B.** (1961) 'Modalities and Intensional Language', *Synthese* 13: 303-22. (Argues that proper names are Millian and that true identity statements involving them are always necessary; referred to in §2.)
- Millikan, R.** (1984) *Language, Thought, and Other Biological Categories*, Cambridge, MA: MIT Press. (The most detailed proposal of a teleological theory of reference; referred to in §8.)
- Mill, J.S.** (1843) *A System of Logic*, London: Longmans, 1867. (Book 1, Chapter 2 proposes his theory of both 'general and singular' names - referred to in §1.)
- Neale, S.** (1990) *Descriptions*, Cambridge, MA: MIT Press. (A thorough defence of the Russellian approach to definite descriptions. Includes an excellent bibliography.)
- Papineau, D.** (1987) *Reality and Representation*, Oxford: Blackwell. (Proposes a teleological theory of reference.)
- Putnam, H.** (1975) *Mind, Language and Reality: Philosophical Papers*, Cambridge: Cambridge University Press, vol. 2, 196-290. (Very influential argument against description theories and for causal theories of natural-kind and other terms; referred to in §§3 and 4.)
- Putnam, H.** (1983) *Realism and Reason: Philosophical Papers*, Cambridge: Cambridge University Press, vol. 3, 1-25. (Presents his model-theoretic argument that reference is indeterminate; referred to in §9.)
- Quine, W.V.** (1960) *Word and Object*, Cambridge, MA: MIT Press. (Section 12 argues for the inscrutability of reference in the context of arguing for the indeterminacy of translation; referred to in §9.)

Reference

- Reichenbach, H.** (1947) *Elements of Logic*, London: Macmillan.(Section 50 proposes a theory of ‘token-reflexive’ words; referred to in §5.)
- Russell, B.** (1905) ‘On Denoting’, *Mind* 14: 479-93.(A classic paper proposing his famous theory of descriptions; referred to in §6)
- Russell, B.** (1911) ‘Knowledge by Acquaintance and Knowledge by Description’, in *Mysticism and Logic*, London: Allen & Unwin, 1917.(A classic source for the description theory of proper names; referred to in §1.)
- Russell, B.** (1912) *The Problems of Philosophy*, London: Oxford University Press, 1959. (Chapter 5 gives a brief and accessible account of Russell’s theory of proper names and descriptions.)
- Russell, B.** (1918) ‘The Philosophy of Logical Atomism’, in R.C. Marsh (ed.) *Logic and Knowledge*, London: Allen & Unwin, 1956.(Discusses the referential properties of many terms, including ‘egocentric particulars’; referred to in §§1, 5 and 6.)
- Searle, J.** (1958) ‘Proper Names’, *Mind* 67: 166-73.(A very clear presentation of a cluster description theory of names; referred to in §1.)
- Strawson, P.F.** (1950) ‘On Referring’, *Mind* 59: 320-44.(A critique of Russell’s theory of descriptions.)
- Strawson, P.F.** (1959) *Individuals: An Essay in Descriptive Metaphysics*, London: Methuen, 180-3, 190-4. (Proposes a cluster description theory of names; referred to in §1.)

Relevance logic and entailment

'Relevance logic' came into being in the late 1950s, inspired by Wilhelm Ackermann, who rejected certain formulas of the form $A \rightarrow B$ on the grounds that 'the truth of A has nothing to do with the question whether there is a logical connection between B and A '.

The central idea of relevance logic is to give an account of logical consequence, or entailment, for which a connection of relevance between premises and conclusion is a necessary condition. In both classical and intuitionistic logic, this condition is missing, as is highlighted by the validity in those logics of the 'spread law', $A \& \sim A \rightarrow B$; a contradiction 'spreads' to every proposition, and simple inconsistency is equivalent to absolute inconsistency. In relevance logic the spread law fails, and the simple inconsistency of a theory (that a set of formulas entails a contradiction) is distinguished from absolute inconsistency (or triviality: that a set of formulas entails every proposition). The programme of relevance logic is to characterize a logic, or a range of logics, satisfying the relevance condition, and to study theories based on such logics, such as relevant arithmetic and relevant set theory.

1 The calculus of entailment

'Relevance logic' (also known as 'relevant logic') was born in the late 1950s, when Alan Ross Anderson and Nuel D. Belnap were inspired by reading a paper by Wilhelm Ackermann (1956). In that paper, 'Begründung einer strengen Implikation' (A Foundation for a Rigorous Implication), Ackermann set out to characterize an implication relation, or connective, $A \rightarrow B$ in which there was always a connection between A and B . 'Rigorous implication', he wrote, 'is intended to express the idea that a logical connection holds between A and B , that the content of B is part of the content of A ' (1956: 113). Formulas which, though valid in classical logic, were rejected by Ackermann on the grounds that 'the truth of A has nothing to do with the question whether there is a logical connection between B and A ' included

$$\begin{array}{ll} A \rightarrow (B \rightarrow A) & A \rightarrow (B \rightarrow A \& B) \\ A \rightarrow (\sim A \rightarrow B) & A \rightarrow ((A \rightarrow B) \rightarrow B) \\ B \rightarrow (A \rightarrow A) & A \rightarrow (A \rightarrow A) \\ A \& \sim A \rightarrow B. & \end{array}$$

Ackermann began his paper by recalling C.I. Lewis' reasons for developing his calculus of strict implication as a calculus expressing a more reasonable notion of 'implication' or 'entailment' than was current at the time (see Lewis 1914). (That the quest for a relevant notion of entailment goes back well before Lewis has been demonstrated more recently by C.J. Martin (1986) in his investigations into twelfth-century logic.) But Ackermann believed Lewis had not gone far enough. Although the first four formulas above fail when ' $\rightarrow$ ' is interpreted as strict implication, the last three still hold. Moreover, taking ' $\Box$ ' to represent 'is necessary', $\Box A \rightarrow (B \rightarrow A)$ and $\Box \sim A \rightarrow (A \rightarrow B)$ were theses of Lewis' system of strict implication. Ackermann added ' $\Box$ ' to his calculus of rigorous implication (via an absurdity constant ' $\perp$ ' and a definition of ' A is impossible' as $A \rightarrow \perp$) and showed that the above so-called 'paradoxes of strict implication' fail when ' $\rightarrow$ ' is taken as rigorous implication.

Belnap reports returning to Yale University from a year in Europe working with Feys on composing a Gentzen system for Ackermann's system, to write up his Ph.D. dissertation, and ingenuously asking Anderson if he knew anyone in the United States who was interested in Ackermann's work. Anderson, feeling equally isolated in his fascination with Ackermann's ideas, having reviewed them in the *Journal of Symbolic Logic*, was delighted to find, as Belnap put it, a 'fellow Strengenite'. Together they started the research programme which led to more than thirty years' creative work on every aspect of a logic of relevance.

How does their work relate to Ackermann's? The initial motivation was the same, and the same goal which had inspired Lewis: to find a satisfactory explication of the concept of entailment, or (formal) logical consequence. Therefore, in the first decade of work, the focus was on the calculus of entailment (E) incorporating aspects of both relevance and necessity. E validates the same theses as Ackermann's system, and differs from it in several apparently minor ways, which in fact are quite significant. First, ' $\Box A$ ' can be defined directly in the basic system

as $(A \rightarrow A) \rightarrow A$, yielding the same theses, so ' $\perp$ ' and the associated apparatus can be removed. ' $\rightarrow$ ' is itself a modal connective (the ' $\rightarrow$ '-fragment of E is a subsystem of the arrow fragment of S4). Second, two of Ackermann's four rules of inference, namely (γ) and (δ) , though unproblematic in the pure logic, are unacceptable when the logic is extended to a substantive theory. Anderson and Belnap showed that they could be dispensed with. Ackermann's rule of inference (γ) says that a proof of B results immediately from proofs of $A \vee B$ and $\sim A$. The point is this: whenever $A \vee B$ and $\sim A$ are provable in the pure logic of E, so too is B , that is, (γ) is an admissible rule of E, and so is redundant. (Conjectured in 1958, the first proof was not found until ten years later.) But in a theory based on E, there might be proofs of $A \vee B$ and $\sim A$ (that is, these formulas might belong to the theory, extending E by non-logical axioms) without there necessarily being a proof of B (or of $\sim B$ - the theory need not be negation-complete). One major task in the research programme of relevance logic became to show the admissibility of (γ) for various E-theories, once E was shown to admit (γ) itself. We will note later the major disappointment in this search, the failure of (γ) for relevant arithmetic.

The problem with the rule (δ) is similar, but less deep and seminal. (δ) says that from $A \rightarrow (B \rightarrow C)$ and a proof of B one can infer $A \rightarrow C$. Again, the proof of B must be logical. Even if $A \rightarrow (B \rightarrow C)$ and B belong to a theory, $A \rightarrow C$ might not. Indeed, what (δ) seems to promise is a permutation, from $A \rightarrow (B \rightarrow C)$ to $B \rightarrow (A \rightarrow C)$, and that is unacceptable in the modal logic E, being a fallacy of modality. (Let A be $B \rightarrow C$ itself, and consider how the permutation would fail in S4 with ' $\rightarrow$ ' as strict implication.) Such a move is legitimate only if B is appropriately modal, which we can capture with the connective ' $\Box$ '. (δ) can therefore be replaced by an axiom,

$$(A \rightarrow (B \rightarrow C)) \rightarrow (\Box B \rightarrow (A \rightarrow C)),$$

or, equivalently, as Anderson and Belnap chose to do in E,

$$(\Box A \ \& \ \Box B) \rightarrow \Box(A \ \& \ B),$$

with ' $\Box$ ' defined as above.

E, the calculus of entailment, had thus been born; its study was to reveal new worlds in logic, and, before long, to shift attention to a range of related logics. (See [Anderson and Belnap 1975, 1992](#).)

2 The calculus of relevant implication

Anderson and Belnap called E the logic of relevance and necessity. However, it is important to understand that the programme they embarked on did not try to give any general characterization of relevance, or even of relevance as a logical notion. What they identified were what they called fallacies of modality and of relevance. Fallacies of modality were first characterized by them for the ' $\rightarrow$ '-fragment as taking as valid any formula of the form $B \rightarrow (A \rightarrow C)$, where B was a propositional variable. Ackermann had suggested the criterion for his full calculus that neither ' $\rightarrow$ ' nor ' $\perp$ ' should belong to B .

Arguably, neither characterization really identifies what a modal fallacy is. For one thing, neither condition easily generalizes beyond the propositional language given. Adding further connectives or quantifiers and the language of predicates, one must generalize the condition, yet that means one needs some understanding of what motivates it. Pottinger tried to capture a notion suitable for E (see [Anderson and Belnap 1975: 348](#)), but again, how it would generalize further is unclear. These are technical criteria, not proper philosophical analyses. Moreover, while separating the modal from the relevance fallacies is a promising idea, a natural consequence is that the modal fallacies should equally be rejected in S4. Yet $B \rightarrow (A \rightarrow C)$ is a thesis of S4 whenever $A \rightarrow C$ is, regardless of the nature of B , even if it is a propositional variable. The answer is that a modal fallacy is essentially a permutation move, promoting B from its secondary position to a primary one: from $A \rightarrow (B \rightarrow C)$ to $B \rightarrow (A \rightarrow C)$. Such a permutation will permit, *inter alia*, validation of formulas $B \rightarrow (A \rightarrow C)$ where B is a propositional variable (or contains neither ' $\rightarrow$ ' nor ' $\perp$ '). Checking that such formulas are not theses is a useful test that a system does not commit any modal fallacies. But it is not an adequate explication of the notion.

Fallacies of relevance were also characterized by Anderson and Belnap only in technical terms, without any wider elaboration of what was really wrong with them. Their formal tests of relevance were variable-sharing and derivational utility. The former again only applies to the propositional fragment. The idea was that $A \rightarrow B$ is valid only if A and B share a propositional variable. Belnap presented matrices which test for this condition. The

other criterion arose from an attempt by Anderson and Belnap to prove a deduction theorem for Ackermann's system, and led to the creation of a (Fitch-style) natural deduction formulation of E. They developed sets of indices which kept track of relevance through the proof. Conditional proof (the analogue in the natural deduction formulation of the deduction theorem as a result about axiomatic formulations) was permissible only when in the proof of B from A , B 's index contained that assigned to the assumption of A . The idea has been fecund in the development of relevance logic, especially when one realizes that the indices play the role of the assumption lists in a Gentzen-style formulation of the logic as a natural deduction calculus of sequents. But it is still a restricted and technical idea: how will it generalize when we extend the calculus to, say, second-order, and outstrip proof theory? In brief, what is really meant here by relevance?

But separating ideas of relevance from those of modality permitted the creation of the system R of relevant implication. This results from E by relaxing the restriction on permutation. The ' $\rightarrow$ ' of R is a non-modal relevant conditional. It was seen as relating to the ' $\rightarrow$ ' of E as material implication relates to strict implication. Yet in this development lay the effective death-knell of E. If one now starts with R and adds an S4-necessity operator ' $\Box$ ', there are theses containing ' $\Box$ ' only in the position $\Box(A \rightarrow B)$ which do not result from theses of E in which $A \rightarrow B$ has been replaced by $\Box(A \rightarrow B)$. Since this discovery in 1973 (see [Anderson and Belnap 1975: 351-2](#)), work on R and its extensions and subsystems has put E in the shade.

3 The systems of relevance logic

With the creation of R, and more so with its eclipse of E, the stage was set for a study of a full range of systems of relevance logic. Immediate neighbours of E and R were the system T of 'ticket entailment' and the 'mingle' systems EM and RM. Ticket entailment picked up an idea of Ryle's; that conditionals are a kind of inference-ticket, licensing the inference from one factual (non-conditional) statement to another. T is a subsystem of E which effectively restricts even the permutation of conditional antecedents which is allowed in E.

The mingle systems extend E and R by the addition of the axiom $A \rightarrow (A \rightarrow A)$. This looks harmless, lacking the obvious irrelevance of $A \rightarrow (B \rightarrow A)$. But adding the mingle axiom to R leads to the validity of $(A \rightarrow B) \vee (B \rightarrow A)$, and we find one of the classic paradoxes of material implication has returned. RM is an interesting source of study; but it is, in Meyer's phrase, a 'semi-relevant logic'.

With hindsight, one can see the study of E and R as the start of research into a range of systems which in the early 1990s came to be called 'substructural logics'. Two insights led to this development. One was recognition of the Curry-Howard correspondence, the treatment of formulas as types (see [Lambda calculus](#)). This reveals that the ' $\rightarrow$ '-fragment of R was being studied even in the 1930s by Church, in his λ -I calculus. In the λ -calculus one studies functions, developing a framework for the general treatment of functional abstraction and application. The reductions and equivalences of functional expressions are governed by combinators, so that, for example, the combinator B asserts that the composition of functions is associative: $B(fg)h$ reduces to $f(gh)$ (see [Combinatory logic §1](#)). Correlating functional abstraction with conditional proof, functional application with *modus ponens*, there is a natural correlation of the definability of λ -terms, or combinators, with the provability of ' $\rightarrow$ '-formulas (which extends to Cartesian product and conjunctions, and so on). R (or at least, its ' $\rightarrow$ '-fragment) is then seen as BCW-logic, the logic of associativity, permutation and contraction, placed in a family with BCK-, BCI- and other logics. BCI-logic, otherwise known as the ' $\rightarrow$ '-fragment of RW (R without contraction) corresponds to Church's λ -I calculus. The relevance logics can then be classed as those representing functions that 'really depend on their arguments', rejecting the combinators such as K , corresponding to the λ -term $\lambda x, y. x$ (see [Combinatory logic §1](#)).

The other development was 'correspondence theory', the study of the correlation between logical principles and constraints on corresponding relations in the model theory. It started in the study of semantics for modal logics, whereby different conditions on accessibility in possible worlds semantics are seen to match various axioms for necessity, ' $\Box$ ', in the logic modelled. The semantics for relevance logic began to be worked out in the late 1960s and early 1970s by Dunn, the Routleys, Meyer, Maximova and Fine, working largely independently. There were two aspects needing special attention. First, a way of invalidating the spread law had to be found, and that seemed to require modelling A and $\sim A$ simultaneously. One solution was the negation-shift operation, pairing each world w with w^* , whereby $\sim A$ is true at w if and only if A is true at w^* . Suitable constraints on ' $*$ ' give the right involute properties to ' $\sim$ '. The other aspect was the relevance ' $\rightarrow$ ' itself, whether for E or R (or related logics).

The approach which has come to dominate uses a three-place relation on worlds, R , whereby $A \rightarrow B$ is true at x if B is true at b whenever A is true at a and $Rxab$. Again, suitable constraints on R give the desired properties of ' $\rightarrow$ '. Fine's version used a pre-ordering of worlds (as in S4) and an operation of fusion, ' $\circ$ ', such that $A \rightarrow B$ is true at x if B is true at $x \circ a$ whenever A is true at a . This operational semantics makes more transparent the correspondence with the combinators. The correspondences thus set up identify a range of logics, the substructural logics (logics with restricted structural rules), which differ essentially only in their structural rules.

This assumes, however, the invariance of the Boolean apparatus of ' $\checkmark$ amp;' and ' $\vee$ ' erected on the underlying calculus of ' $\sim$ ', ' $\rightarrow$ ', and perhaps fusion ' $\circ$ ', with $A \circ B$ true at $a \circ b$ whenever A is true at a and B at b . ' $\circ$ ' is a 'relevant conjunction', $A \circ B \rightarrow C$ being equivalent to $A \rightarrow (B \rightarrow C)$. ' $\checkmark$ amp;' and ' $\vee$ ' need not be so invariant, however. Ackermann noted that distribution of ' $\checkmark$ amp;' over ' $\vee$ ' no longer followed from the other axioms when restricted to yield rigorous implication, and so added it as a separate axiom. Anderson and Belnap followed him, supporting the case by yet another criterion of relevance, applied to first-degree entailments ($A \rightarrow B$ where A and B contain only ' $\sim$ ', ' $\checkmark$ amp;' and ' $\vee$ '); that they be tautological: putting A in disjunctive normal form (DNF) and B in conjunctive (CNF), every disjunct on the left shares an atom with every conjunct on the right, roughly, B is 'contained in' A . But some have argued that dropping the combinator K ('weakening') as the source of irrelevance in full classical logic, should yield a non-distributive calculus. Interestingly, every classical tautology (in ' $\sim$ ', ' $\checkmark$ amp;' and ' $\vee$ ') is provable in R . But this fact fails for the non-distributive calculus. For example, the axiom 'sum' of *Principia Mathematica*,

$$(Q \supset R) \supset ((P \vee Q) \supset (P \vee R)),$$

with ' $A \supset B$ ' for ' $\sim A \vee B$ ', is unprovable in LR (R without distribution). LR was the focus of much research in automated theorem proving in the 1980s.

Dropping W ('contraction') from LR yields a fragment of Girard's linear logic (1987), as does dropping distribution from RW (see [Linear logic](#)). Girard's rich theory, encompassing additive connectives (' $\checkmark$ amp;' and ' $\vee$ ' - there written ' $\checkmark$ amp;' and ' $\oplus$ '), multiplicatives (' $\rightarrow$ ' and ' $\circ$ ' - there ' $\multimap$ ' and ' $\otimes$ ') and exponentials (the modal operators '!' and '?'), tracking the essential and profligate use of resources, picks up Church's original interest in BCI and has played a central role recently in theoretical computer science.

Those relevance logics (Ackermann's system is one exception) which reject the spread law in the form $A \& \sim A \vdash B$, are paraconsistent logics (see [Paraconsistent logic](#)). For some, they even share a motivation, namely, to allow logical treatment of inconsistent theories without making them trivial (their deductive closure does not embrace everything, as in classical logic). But for others, their motivation is different: many paraconsistent logics arise from an interest in dialetheism, the belief that some inconsistent theories are true. Wider still run the 'sociative logics' (a term of Sylvan's), encompassing relevance, connexive, non-transitive and a whole range of, as he puts it, 'broadly relevant' logics.

4 Relevant predication and arithmetic

This survey of relevance logics suggests an undue concentration on the zero-order (propositional) fragment, and that reflects the facts. But important results have been established for first-order relevance logic, relevant arithmetic and so on. For too long, perhaps, it was thought that the extension to first-order was straightforward. Fine's discovery that plausible ways of extending the proof theory and semantics to first-order quantifiers led to incompleteness was a surprise, not really softened by his later elaboration of a semantics which did not validate the unprovable formulas (see [Anderson and Belnap 1992: §§52-3](#)). Urquhart's demonstration of undecidability for propositional T, E and R settled an open problem of twenty years' standing, again in a surprising way (1992: §65). (The non-distributive logics are decidable.) The failure of (γ) for the formulation $R^\#$ of relevant arithmetic was another surprise, blocking a natural line of research.

The exposure of these negative results in the 1980s was not unlike the exposure of the limitative results for classical mathematics (incommensurability, uncountability, incompleteness). At first, relevance logic lost momentum as they were settled, and settled in such surprising and negative ways. Nonetheless, limitative results notoriously give rise, once assimilated, to new avenues of research. They are appearing in relevance logic, two notable ones being the attempt to characterize a notion of relevant predication by Dunn and others ([Anderson and Belnap 1992: §74](#)) - recognizing that the extension to first-order is not straightforward; and the search for a

relevant arithmetic which admits (γ) (1992: §72) - perhaps by strengthening the rule of induction.

See also: [Consequence, conceptions of](#); [Logical and mathematical terms, glossary of](#); [Logical laws](#); [Modal logic](#)
STEPHEN READ

References and further reading

- Ackermann, W.** (1956) 'Begründung einer strengen Implikation' (A Foundation for a Rigorous Implication), *Journal of Symbolic Logic* 21: 113-28. (The original paper which started it all.)
- Anderson, A.R. and Belnap, N.D.** (eds) (1975, 1992) *Entailment: The Logic of Relevance and Necessity*, Princeton, NJ: Princeton University Press, 2 vols. (The classic source for their work and others'; includes an extensive bibliography.)
- Došen, K. and Schroeder-Heister, P.** (eds) (1993) *Substructural Logics*, Oxford: Clarendon Press. (Places relevance logic in the framework of wider study of logics with restricted structural rules.)
- Dunn, J.M.** (1986) 'Relevance Logic and Entailment', in D. Gabbay and F. Guenther (eds) *Handbook of Philosophical Logic*, vol. 3, *Alternatives to Classical Logic*, Dordrecht: Reidel, 117-229. (A very useful technical survey.)
- Friedman, H. and Meyer, R.K.** (1992) 'Whither Relevant Arithmetic?', *Journal of Symbolic Logic* 57: 824-31. (Shows the failure of (γ) for relevant arithmetic.)
- Girard, J.-Y.** (1987) 'Linear Logic', *Theoretical Computer Science* 50: 1-102. (The first and definitive presentation of linear logic.)
- Lewis, C.I.** (1914) 'The Calculus of Strict Implication', *Mind* 23: 240-7. (One of Lewis' early papers in which he described his project of formalizing a logic of entailment.)
- Martin, C.J.** (1986) 'William's Machine', *Journal of Philosophy* 83: 564-72. (The debate over relevance in twelfth century logic.)
- Read, S.** (1988) *Relevant Logic*, Oxford: Blackwell. (An elementary presentation of the philosophical basis of relevance logic and a formal system for a range of relevance logics.)
- Sylvan, R. [née Routley] with Plumwood, V., Meyer, R.K. and Brady, R.** (1982) *Relevant Logics and Their Rivals*, vol. 1, *The Basic Philosophical and Semantical Theory*, Atascadero, CA: Ridgeview. (Places zero-order relevance logic in a setting of non-classical intensional logics.)
- Sylvan, R. and Norman, J.** (eds) (1989) *Directions in Relevant Logic*, Dordrecht: Kluwer. (A collection of papers including those given at a memorial conference for Alan Anderson in 1974, some updated and others additional.)
- Thistlewaite, P.B., McRobbie, M.A. and Meyer, R.K.** (1988) *Automated Theorem-Proving in Non-Classical Logics*, London: Pitman. (The computer program KRIPKE, using a search tree based on models for non-distributive relevance logic.)

Rhetoric

Rhetoric is the power to persuade, especially about political or public affairs. Sometimes philosophy has defined itself in opposition to rhetoric - Plato invented the term 'rhetoric' so that philosophy could define itself by contrast, and distinctions like that between persuasion and knowledge have been popular ever since. Sometimes philosophy has used rhetorical techniques or materials to advance its own projects. Some of its techniques, especially topics of invention, the classification of issues, and tropes or figures of speech, are occasionally employed by philosophers. The philosophical question is whether these techniques have any interest beyond efficacy. What is the relation between techniques effective in persuading others and methods for making up one's own mind? Is there any connection between the most persuasive case and the best decision? Is there a relation between the judgments of appropriateness and decorum exercised by the rhetorician, and the judgments of appropriateness exercised by the person of practical wisdom? Do judgments about probability, ambiguity and uncertainty, and judgments under constraints of time or the need for decision, aspire to the ideal of perfect rationality, to which they are doomed to fall short, or do these kinds of judgment have an integrity of their own? Apart from supplying useful techniques, an art of persuasion also raises philosophic questions concerning the relation between rhetoric and logic, rhetoric and ethics, and rhetoric and poetics.

1 Philosophy and rhetoric at their beginning

Plato invented the term 'rhetoric' as a contrast term against which philosophy could define itself. Philosophy and rhetoric both proposed new truths and apparently powerful methods that threatened existing moral codes and authorities. As modes of empowerment, self-making and self-consciousness they could as easily become enemies as allies. Thus while philosophy attempted to achieve self-consciousness and power through an awareness of thought and being, rhetoric focused on an awareness of language and the circumstances of speaking and acting. Cicero interpreted the Platonic separation of philosophy from rhetoric as that of Socrates dividing wisdom from eloquence, and saw that the task of philosophical rhetoric was to overcome this separation. Prior to Socrates, he says in *de Oratore*, 'the same teaching seems to have imparted education both in right conduct and in good speech...the same masters gave instruction in both ethics and rhetoric' (see *Cicero* §§1, 2). From the beginning, then, rhetoric already had the pejorative connotation it continues to have, in spite of periodic attempts to speak in its defence.

Rhetoric's importance for philosophy seems greatest in periods of political and cultural confrontation and pluralism, for example, during Roman expansion, the Renaissance and today. These are all times of 'linguistic turns' in philosophy and, in addition, of fruitful interactions between philosophic and legal argumentation and interpretation. Linguistic, legal and pragmatic turns are rhetorical phenomena that direct attention to effective reasoned communication. In such times, the revival of rhetoric is motivated by a sense that philosophy itself has become too professionalized and too remote from human concerns. Rhetoric is then invoked, as in Cicero, to bring philosophy back down to earth. That neither Rome nor the Renaissance is considered a golden age of philosophy is probably a reflection of their rhetorical character.

Whether or not the history of philosophy is a series of footnotes to Plato, the history of the philosophical issues of rhetoric certainly can be read that way. The *Gorgias* begins by showing the difficulty in defining rhetoric. Gorgias finds it hard not only to define rhetoric logically, but to offer a definition in the more serious sense of keeping it within its proper sphere. In Socrates' hands the universality of rhetoric becomes a sign of emptiness, not power, and its flexibility turns from resourcefulness into slavishness. The conversation between Socrates and Polus makes this philosophy of rhetoric the first example of 'applied ethics', in which one searches for ethical constraints on an activity. The conversation between Socrates and Callicles originates the tradition of 'professional ethics' which relates the internal norms of a practice to the sort of person the practitioner is.

Other Platonic dialogues amplify these considerations and raise further questions. The same rhetoric that appears in the *Gorgias* as a power of domination appears in the *Protagoras* in a context of mutual agreement. The most obvious use of rhetoric - arguing both sides of a question - can be either to exercise power unilaterally or to provide the basis of community, just as competitive Homeric virtues either underlie or are transformed into the cooperative virtues of the polis. In the twentieth century, Kenneth Burke (1950) highlights this shift by changing

the central occupation of rhetoric from persuasion to ‘identification’. The relation between the competitive and cooperative facets of rhetorical argument sets another philosophic problem for rhetoric: Under what conditions can the power to argue both sides of a question become the power to uncover truth through the free competition or marketplace of ideas? (In *On Liberty*, Mill develops his vision of truth emerging through controversy by reference to Cicero, ‘the second greatest orator in antiquity’.) The introductory conversation in the *Protagoras* between Hippocrates and Socrates raises questions about the relation between learning from the sophists and being a sophist oneself. The dialogue then presents Socrates in conversation with the sophists Protagoras, Hippias and Prodicus. Protagoras limits the domain of rhetoric to political questions, justice and civic virtue, while Hippias would expand it to cover the natural sciences, and Prodicus would extend it to literary interpretation. Whether rhetoric is restricted to political questions, whether there is a ‘rhetoric of science’ or further uses of rhetoric outside politics all become standard questions for rhetoric from then onwards.

The *Phaedrus* turns to a third possible alliance which defines rhetoric in relation to poetics. In addition to rhetoric and logic (what does the rhetorician know?) and rhetoric and ethics (what are the uses and abuses of this power, and what sort of person does one become in practising rhetoric?) there is the relation between rhetoric and poetry, another activity that either employs charms not available to more rational approaches or operates at a further remove from reality. The *Phaedrus* raises further questions concerning the relation between written and spoken language, and so between the production and reception of discourse, a line of investigation that eventually leads from rhetoric to hermeneutics. Questions from the *Gorgias* concerning rhetoric and knowledge are formulated in the *Phaedrus* in a pointed manner as the relation between rhetoric and sincerity and deception, or between persuasion, *eros*, beauty, and seduction. Burke’s investigations of the relations between serious and playful uses of language are a modern rediscovery of this set of questions, as are the recent revivals of rhetoric in *de Man* and *Derrida*.

The *Phaedrus* ends with Socrates asking whether there is a rhetoric for philosophy - that is, whether there are modes of communication that are especially suited for philosophy. But Socrates’ questions about written versus spoken language, and long speeches versus questions and answers, at the same time pose the parallel problem of whether rhetoric has a philosophy, whether it has ethical or metaphysical commitments. The answer is parallel to that for the analogous question about mathematics: does the mathematical enterprise inherently have a ‘philosophy’. There seems to be a natural affinity between mathematics and one type of philosophy, namely Platonism. On the other hand there are as many philosophies of mathematics as there are types of philosophy. Similarly, there are affinities of rhetoric with scepticism and relativism, and there are as many philosophies of rhetoric as there are kinds of philosophy (see *Plato* §§4, 5).

2 Philosophical problems posed by rhetoric

Rhetoric is persuasive communication, especially about practice and politics. Whether rhetoric is a subject for philosophy depends on whether persuasion raises issues of its own apart from those exhausted by logic, ethics and psychology. If the way people were persuaded had nothing at all to do with logical cogency or ethical responsibility, or if it was simply reducible to one of them, there would be nothing philosophic about rhetoric. It is only because there seems to be some connection between logical cogency and rhetorical persuasiveness, between the trust accorded a persuasive speaker and the confidence we have in good ethical agents, and between figurative language and something more than mere ornament, that rhetoric is a subject for philosophy. While rhetorical theory has had negligible influence on the history of philosophy, the problems raised by rhetorical practice are significant philosophical problems.

All the issues raised by Plato’s dialogues could have become philosophical problems in a variety of ways, but they were disciplined by Aristotle’s organization of knowledge. The details of Aristotle’s *Rhetoric* have had little impact on the history of either philosophy or rhetoric, but his placement of rhetoric relative to other arts and sciences has organized the trajectory of the relation of philosophy and rhetoric. Rhetoric, he says, is the offshoot of dialectic and politics. There are three ends and kinds of rhetoric: political or deliberative rhetoric directed towards utility; judicial or forensic rhetoric which determines justice and injustice; and epideictic or demonstrative rhetoric which concerns worthiness and blameworthiness. There are three sources of persuasion or belief (*pistis*) in rhetoric - the character (*êthos*) of the speaker, the passions of the audience, and the ‘speech (*logos*) itself, in so far as it proves or seems to prove’ - generating the trio implicit in Plato: rhetoric and logic, rhetoric and ethics, rhetoric and

poetics.

The meaning and scope of all these terms of opposition (logic, ethics and poetics) shifts as the domain of rhetoric itself expands and contracts. Rhetoric is potentially about everything that can be communicated, and so about all thought and language. But the use of speech in deliberation about the concrete and practical particular is always at its core. As ethics after Aristotle became less political, so too did rhetoric. In Plato's dialogue, Protagoras defines his subject politically as the abilities without which men could not be citizens. Others softened the definition to the ability to speak on those matters on which people would be ashamed not to have an opinion, that is, the field of common sense and common understanding. Austin's (1962) inquiry captures part of the idea of persuasive communication apart from a political context, and shows the alliance between the linguistic and the rhetorical turns of philosophy (see Austin, J.L.). Some argue that rhetoric is as broad as human speech in general, with extensions to general rhetorics like the 'rhetoric of inquiry' or 'rhetoric of the human sciences' or, as with Burke, to all strategies for encompassing situations. Others would restrict it to public or political affairs. The broadening of rhetoric is sometimes a reflection of the lack of political, or even practical, functioning for rhetoric, and sometimes a corrective to the impractical specialization and professionalization of rhetoric itself. The logic, ethics and poetics against which rhetoric defines itself are usually much broader than the three sources of proof in Aristotle's account.

3 Rhetorical methods and philosophy

As rhetoric developed, it elaborated specific techniques for persuasion, and even if rhetorical activity had no philosophic dimensions, the history of rhetoric would still offer a usable history of persuasion. The first rhetorical method with philosophic import is that of topical invention, which brings into focus the relation between rhetoric and logic. *Topos* is the Greek word for place, translated into Latin as *locus*. *Koinoi topoi* became *loci communi* and then, in English, 'commonplaces,' places to find arguments, pigeon-holes for classifying appeals, or major premises from which to derive particular conclusions. Topics are the means for finding something persuasive to say in a practical situation. Aristotle defines rhetoric as an art of finding in any case the available means of persuasion. Subsequent rhetoricians make invention (*heuresis*, *inventio*) the first of five parts of rhetoric, alongside judgment (sometimes called disposition or arrangement), style, memory and delivery. Any of these can be a merely verbal technique or an art with philosophical significance. The topics can range from the logical ('similar effects have similar causes') to more substantive considerations ('follow the money'; '*cherchez la femme*'; 'we all know that Athenians are acquisitive'). Often there is a connection between methods for finding arguments and systems of artificial memory for recalling them. When the rhetorical invention of arguments is replaced by a scientific discovery of things, only style remains in rhetoric's domain. Ramus' (1574) division of labour between a topical dialectic or logic and a rhetoric confined to style and arrangement in the early modern period is perhaps the most historically influential of such moves.

Scientific method might supplant rhetorical invention, but the methods of scientific discovery which replace rhetorical invention were themselves derived from topical invention. Much of the meaning of necessity in seventeenth-century science, for example, derives from pleas of necessity as an excuse or justification in legal proceedings. Stephen Toulmin's (1964) 'inference warrants', which supply the connection between grounds and claims, is one recent reappearance of the topics; Chaim Perelman's (1969) *loci* for establishing connections is another.

Hume finds the meaning of 'personal merit' through epideictic rhetoric by 'displaying the praises of any humane, beneficent man'. He refers to the considerations used as the 'topics of praise'. The inter-relations among the three kinds of rhetoric, and the kinds of argument appropriate to each, generate significant philosophical problems. Why, as Hume asks, does utility please? What is the relation between the impartial spectator whose judgment is the focus of epideictic rhetoric, the judge whose verdicts are the end of forensic rhetoric, and the ideal deliberative agent? What is the relation between the right and the good, or between what is best and what should be done? What is the connection between action-guiding and agent-evaluating reasons?

The other two philosophically interesting rhetorical methods besides topical invention can trace their ancestry back to Plato and Aristotle, but were really developed in later rhetorics which emphasized judicial oratory and the performance values associated primarily with demonstrative rhetoric. Issues - *staseis* in Greek and *constitutiones* in Latin - were rhetoric's central technique for most of its history. The classification of issues evolved out of

practices of judicial rhetoric, which is always the most systematic of the three kinds of rhetoric, and they draw attention to the relation between rhetoric and ethics. There are, usually, four *staseis*: (1) the question of fact or conjectural issue; (2) the definitive stasis; (3) the qualitative issue concerning mitigating or aggravating circumstances; and (4) the translative issue of whether a given court is the appropriate forum for judgment. Kant, not long after Hume, explains the ‘principles of any transcendental deduction’ by referring to stasis theory: ‘Jurists, when speaking of rights and claims, distinguish in a legal action the question of right (*quid juris*) from the question of fact (*quid facti*)’ (1787: B116-7). His question of right is the translative issue, by what right knowledge can be judged to have a ‘legal title’, so the foundations of knowledge cannot be secured by pointing to facts about the success of science. Austin’s ‘plea for excuses’ and contemporary arguments about responsibility invite reconsideration of the rich problems that revolve around the qualitative issue.

If the topics emphasize the relations between rhetoric and logic, and the theory of issues or *staseis* develops the relation between rhetoric and ethics, the third rhetorical method, that of figures or tropes, concerns the relation of rhetoric to poetics. Augustine, in the *de Doctrina Christiana*, used rhetoric to teach Christians to read their Bible figuratively and to adjust contradictory passages by considering the circumstances to which speakers and authors adapted their statements and meanings. For Vico (1744), a near contemporary of Hume and Kant, the ‘tropes are corollaries of poetic logic’. Borrowing from a formulation that seems to originate with Ramus, Vico names four primary tropes: metaphor, metonymy, synecdoche and irony. These tropes are not merely elegant variations on things that could be expressed more directly, but reveal original truths hidden by more sophisticated and professionalized philosophy. In the hands of our own contemporaries such as Kenneth Burke (1950) and Hayden White (1973), the figures move from methods of description or redescription into methods of seeing and encompassing situations, as rhetoric is expanded from a verbal to a universal art. Perelman, similarly, constructs a relation between argumentative forms and figures of speech and thought. Vico has become a hero to those who think that progress will come in philosophy from being anti-Cartesian, and championing conversation and communication over more limited, anti-rhetorical, forms of rationality. These examples show how even the most language-centred or poetic dimension of rhetoric has an ethical agenda. The rhetorical emphasis on saying what is appropriate provides the connection between ethical and poetic demands for timeliness and decorum. The current revival and popularity of rhetoric in philosophy often takes the form of a programme to replace science by conversation as the model for community and rationality and the hope for democracy.

See also: [Language, medieval theories of](#); [Language, Renaissance philosophy of](#); [Legal reasoning and interpretation](#)

EUGENE GARVER

References and further reading

- Aristotle** (c. mid 4th century BC) *On Rhetoric: A Theory of Civic Discourse*, trans. with introduction, notes and appendices by G.A. Kennedy, New York and Oxford: Oxford University Press, 1991.(The most readable and reliable English translation available.)
- Augustine** (c.396-427) *De Doctrina Christiana (On Christian Doctrine)*, trans. D.W. Robertson, Jr, Indianapolis, IN: Bobbs-Merrill, 1958.(The most influential work explicitly adapting rhetoric to Christian religious and theological purposes.)
- Austin, J.L.** (1962) *How to do things with words*, Cambridge, MA: Harvard University Press.(A recovery of the rhetorical dimension of language under the title ‘performative utterance’.)
- Barthes, R.** (1970) ‘L’Ancienne Rhétorique: Aide-mémoire’, *Communications* 16 (1970): 172-229; ‘The Old Rhetoric: An Aide-Mémoire’, in *The Semiotic Challenge*, trans. R. Howard, New York: Hill & Wang, 1988, 11-94.(Designed as a summary of the history of rhetoric for students of contemporary literary criticism.)
- Burke, K.** (1950) *A Rhetoric of Motives*, New York: Prentice Hall.(A cataloguing of the formal techniques for persuasion.)
- Burke, K.** (1965) *Permanence and Change: An Anatomy of Purpose*, Indianapolis, IN: Bobbs-Merill.(An analysis of uses of poetry to solve ethical conflicts through rhetoric as ‘an art of appeal’.)
- Cicero** (55 BC) *de Oratore*, trans. E.W. Sutton and H. Rackham, London: Loeb Classical Library, 1959.(Cicero’s most philosophical treatment of rhetoric. Not a practical handbook but a dialogue on the place of rhetoric in ethical and political life.)
- Conley, T.M.** (1990) *Rhetoric in the European Tradition*, New York and London: Longman; Chicago, IL: University of Chicago Press, 1994.(A textbook, full of summaries and bibliographies; the best one-volume

history of rhetoric.)

- Garver, E.** (1994) *Aristotle's Rhetoric: An Art of Character*, Chicago, IL: University of Chicago Press.(A philosophical analysis of Aristotle's Rhetoric.)
- Hume, D.** (1748, 1751) *Enquiries Concerning the Human Understanding and Concerning the Principles of Morals*, ed. L.A. Selby-Bigge, Oxford: Clarendon Press, 1966.(A moral theory based on the inference patterns of epideictic rhetoric.)
- Kant, I.** (1787) *Critique of Pure Reason*, trans. N. Kemp Smith, London: Macmillan, 1929.(A critical philosophy rooted in justificatory rhetoric.)
- Kennedy, G.** (1980) *Classical rhetoric and its Christian and secular tradition from ancient to modern times*, Chapel Hill, NC: University of North Carolina Press.(One of a series of informative histories of rhetoric by this author.)
- McKeon, R.** (1987) *Rhetoric: Essays in Invention and Discovery*, Woodbridge, CT: Oxbow Press.(Contains the influential 'Rhetoric in the Middle Ages' and other articles by an important but difficult historian and philosopher of rhetoric.)
- Mill, J.S.** (1859) *On Liberty*, New York: Bobbs-Merill, 1956.(Chapter 2, in particular, concerns the power of controversy and discussion to reveal truth.)
- Perelman, C. and Olbrecht-Tyteca, L.** (1969) *The New Rhetoric: A Treatise on Argumentation*, Notre Dame, IN: University of Notre Dame Press.(A treatment of the argumentative force of figures of speech and the topics or *loci communes*.)
- Plato** (c.390-347BC) *The Collected Dialogues of Plato*, ed. E. Hamilton and H. Cairns, Princeton, NJ: Princeton University Press, 1971.(All Platos's dialogues collected in one volume, with an introduction and prefatory notes.)
- Ramus, P.** (1574) *Logikē*, trans. M. Roll, London: Scholar Press, 1966.(The only edition available in English of this highly influential work.)
- Richards, I.A.** (1936) *The Philosophy of Rhetoric*, New York and London: Oxford University Press.(Proposes a new rhetoric that studies misunderstandings and their remedies.)
- Ricoeur, P.** (1977) *The Rule of Metaphor*, Toronto, Ont.: University of Toronto Press. (A historical and philosophical investigation of the interrelations between rhetorical and poetic approaches to language and literature.)
- Trimpf, W.** (1983) *Muses of One Mind: The Literary Analysis of Experience and Its Continuity*, Princeton, NJ: Princeton University Press.(Traces the competition and joint contributions of philosophy and rhetoric to literary theory, from the Greeks to the Renaissance.)
- Toulmin, S.** (1964) *The Uses of Argument*, Cambridge: Cambridge University Press. (A re-discovery, under the name 'inference warrants', of the topics of classical rhetoric.)
- Vico, G.** (1744) *The New Science*, trans. T. Bergin and M. Fish, Ithaca, NY: Cornell University Press, 1984.(A defence of rhetoric against philosophy, and in particular the epistemic and ethical value of figurative over literal language.)
- White, H.** (1973) *Metahistory: The Historical Imagination in Nineteenth-Century Europe*, Baltimore, MD and London: Johns Hopkins University Press.(Uses of Kenneth Burke's theory of tropes to analyse historical narratives and philosophies of history.)

Scope

Scope is a notion used by logicians and linguists in describing artificial and natural languages. It is best introduced in terms of the languages of formal logic. Consider a particular occurrence of an operator in a sentence - say, that of ' $\rightarrow$ ' in (1) below, or that of the universal quantifier ' $\forall$ ' in (2) below.

(1) $A \rightarrow (B \& C)$

(2) $\forall x(Bxy \rightarrow \exists yAxy)$

Speaking intuitively, the scope of the operator is that part of the sentence which it governs. The scope of ' $\rightarrow$ ' in (1) is the whole sentence; this renders the whole sentence a conditional. The scope of '*Čamp;*', on the other hand, is just ' $(B \& C)$ '. In (2), the scope of the quantifier ' $\forall$ ' is the whole sentence, which allows it to bind every occurrence of x . The scope of ' $\exists$ ' is only ' $\exists yAxy$ '. Since ' Bxy ' is outside its scope, the ' y ' in ' Bxy ' is left unbound.

Although the importance of scope is semantic, it is usually identified with a syntactic relation between an occurrence of an expression in a sentence and a part of that sentence. In sentence and quantifier logic, the scope of an occurrence of an operator in a sentence S is usually defined as the smallest sub-sentence of S containing the occurrence. This ensures that for any two (occurrences of) operators with overlapping scope, exactly one is in the other's scope. In ' $A \rightarrow (B \& C)$ ', for example, '*Čamp;*' is in the scope of ' $\rightarrow$ ', and not vice versa. When the scope of one operator includes that of a second, the first is said to have wider scope than the second; the second's scope is narrower than that of the first.

Many semantically significant properties are characterized in terms of scope. The main logical operator of a sentence is the operator with the whole sentence as its scope. Thus conjunction, universal quantification, necessitation and so on are defined in terms of scope. A variable is bound whenever it occurs in the scope of a quantifier on it. A *de re* sentence is a sentence, such as ' $\exists x$ I believe that x is a spy', in which a variable in the scope of a non-extensional operator (like 'believes' or 'necessarily') is bound by a quantifier outside that operator's scope. *De dicto* sentences are ones which do not involve such 'quantifying in' (see [De re/de dicto](#)).

Differences in relative scope can affect a sentence's truth-conditions. 'Many hate a few' has an interpretation, in which 'many' gets wide scope, on which it says that many have the property 'hating a few'. It has a quite different reading in which 'a few' gets wide scope, on which it says that a few have the property 'being hated by many'. If the quantifier 'someone' in 'I believe that someone is a spy' has a narrower scope than 'believes', the sentence does not imply that there is some individual whom I think is a spy; if the quantifier's scope is wider, it does imply this.

In a well-designed formal language, the relative scope of operators is indicated by surface syntax, but this is not inevitably true in natural language. A natural language sentence with n operators has potentially $n!$ (n factorial: $n(n-1) \dots 3 \times 2 \times 1$) readings, since there are in principle that many ways of ordering the scopes of the operators. (Not all such orderings need be linguistically possible, of course.) Such ambiguities need to be resolved by an interpretive process, or a user's intentions, before the sentence can be evaluated for truth. Some syntactic theories posit a stage in sentence generation ('logical form') at which quantifier scope ambiguities are resolved by moving quantifiers so that their relative positions encode their scope (see [Logical form](#)).

A quantifier's scope should be the domain in which pronouns 'anaphoric on' (referring back to) the quantifier behave as variables bound by it. For example, the occurrence of 'her' in

(1) Every woman knew a man who loved her,

if anaphoric on 'every woman', behaves as a bound variable, since (1) is then understood as equivalent to

(1') $\forall x(x \text{ is a woman} \rightarrow x \text{ knew a man who loved } x)$,

and the scope of 'every woman' is therefore the whole sentence.

Observe that only some uses of pronouns anaphoric on quantifiers behave like variables bound by those

quantifiers. For example, consider the sentence

(2) Garth bought some dogs and Wayne vaccinated them.

Even when anaphoric on ‘some dogs’, ‘them’ in (2) cannot be treated as a bound variable. If it were a bound variable, then the sentence as a whole would be equivalent to

(3) There are some dogs Garth bought which Wayne vaccinated,

but (3) does not imply that Wayne vaccinated *all* the relevant dogs, as (2) does. (Gareth Evans (1985) seems to be the first person to have noticed this.) So ‘them’ is to be excluded from the scope of the quantifier ‘some dogs’. In accounts of natural language grammar which identify sentences with tree structures, quantifier scope is often identified with a structural relation between tree parts called c-command. While an expression in one sentence may be anaphoric on one in another, it is not possible for an expression in one sentence to c-command an expression in another. In particular, ‘some dogs’ in (2) does not c-command ‘them’. It follows, on such accounts, that not all pronouns anaphoric on a quantifier are bound thereby.

Russell (1910-13) speaks of the scope of a description in giving his ‘definition in use’ for (that is, method for eliminating) definite descriptions. When the scope of a description ‘the *F* in a sentence *S* is a sub-sentence ‘*T*(the *F*)’ of *S*, Russell says *S* is equivalent to the result of substituting (a formalization of) ‘there is exactly one *F*, and *T*(it)’ for ‘*T*(the *F*)’ in *S*. For example, if, in ‘Possibly, the mayor is mad’, the scope of the description is the whole sentence, the sentence is equivalent to ‘There is exactly one mayor and possibly it is mad’. If the scope of the description is just ‘The mayor is mad’, the sentence is equivalent to ‘Possibly: there is exactly one mayor and it is mad’. Russell’s account makes the definite article semantically analogous to a binary quantifier.

Expressions other than quantifiers and connectives also have scope. Variable binders (such as function abstractors) and verbs which take sentential complements (for example, ‘to believe’) are examples. Standard definitions extend straightforwardly here.

One hears it said that proper names ‘always take wide scope’. This seems intended as the claim that what is said by a use of a sentence *S*(*n*), with *n* a proper name, is what would have been said by an appropriately related use of a sentence of the form ‘*n* is such that *S*(it)’, with ‘it’ anaphoric on *n*.

See also: [Anaphora](#); [Descriptions](#); [Quantifiers, substitutional and objectual](#)

MARK RICHARD

References and further reading

- Evans, G.** (1985) *Collected Papers*, Oxford: Oxford University Press. (Essays 4, 5 and 8 discuss the interpretation of pronouns as bound variables. Somewhat difficult.)
- Higginbotham, J.** (1980) ‘Pronouns and Bound Variables’, *Linguistic Inquiry* 14: 679-708. (Proposes identifying Russell’s notion of scope with c-command. Difficult.)
- Neale, S.** (1990) *Descriptions*, Cambridge, MA: MIT Press. (Discusses and defends Russell’s account of descriptions. Also discusses contemporary accounts of binding and anaphoric relations. A good introduction to these topics.)
- Russell, B.** (1905) ‘On Denoting’, *Mind* 14: 479-93. (Russell’s first presentation of his account of definite descriptions.)
- Russell, B. and Whitehead, A.N.** (1910-13) *Principia Mathematica*, Cambridge: Cambridge University Press, 3 vols; 2nd edn, 1925-7. (Volume 1 includes a formal treatment of descriptions and the scopes thereof.)

Second- and higher-order logics

In first-order predicate logic there are symbols for fixed individuals, relations and functions on a given universe of individuals and there are variables ranging over the individuals, with associated quantifiers. Second-order logic adds variables ranging over relations and functions on the universe of individuals, and associated quantifiers, which are called second-order variables and quantifiers. Sometimes one also adds symbols for fixed higher-order relations and functions among and on the relations, functions and individuals of the original universe. One can add third-order variables ranging over relations and functions among and on the relations, functions and individuals on the universe, with associated quantifiers, and so on, to yield logics of even higher order. It is usual to use proof systems for higher-order logics (that is, logics beyond first-order) that include analogues of the first-order quantifier rules for all quantifiers.

An extensional n -ary relation variable in effect ranges over arbitrary sets of n -tuples of members of the universe. (Functions are omitted here for simplicity: remarks about them parallel those for relations.) If the set of sets of n -tuples of members of a universe is fully determined once the universe itself is given, then the truth-values of sentences involving second-order quantifiers are determined in a structure like the ones used for first-order logic. However, if the notion of the set of all sets of n -tuples of members of a universe is specified in terms of some theory about sets or relations, then the universe of a structure must be supplemented by specifications of the domains of the various higher-order variables. No matter what theory one adopts, there are infinitely many choices for such domains compatible with the theory over any infinite universe. This casts doubt on the apparent clarity of the notion of ‘all n -ary relations on a domain’: since the notion cannot be defined categorically in terms of the domain using any theory whatsoever, how could it be well-determined?

1 Standard second-order logic

Only second-order quantification over relations will be discussed here. Higher-order quantification and quantification over functions can be treated in an entirely parallel way. Symbols for higher-order relations and functions will be omitted for simplicity. Supplement the language of first-order predicate logic (see **Predicate calculus**) with variables X_n^m for m -ary relations, for all natural numbers m and n . Add to the definition of a formula of first-order predicate logic that, for all natural numbers m and n , $X_n^m t_1 \dots t_m$ is an atomic formula, where each t_i is a constant symbol or first-order variable; and that if ϕ is a formula, then so are $(\forall X_n^m)\phi$ and $(\exists X_n^m)\phi$. Free and bound second-order variables are defined in exact parallel to the definitions for first-order variables, and a sentence is a formula with no free variables of any order.

It will be convenient always to require of any given language that there be infinitely many relation symbols of each possible number of places and infinitely many constant symbols not in it. The definition of truth in a structure for first-order predicate logic is extended to the *standard semantics* for second-order logic by supplying additional clauses for second-order quantification as follows. Consider a structure A with language $\mathcal{L}$. Let P be some definite m -placed relation symbol not in $\mathcal{L}$, and define $\psi(P)$ to be the sentence which results from replacing every free occurrence of X_n^m in ψ by P .

- (1) If the sentence ϕ is of the form $(\forall X_n^m)\psi$ for some m and n , then the truth-value of ϕ is T if the truth-value of $\psi(P)$ is T in every expansion of A to a structure the language that includes every member of $\mathcal{L}$ and also P . It is F otherwise.
- (2) If the sentence ϕ is of the form $(\exists X_n^m)\psi$ for some m and n , then the truth-value of ϕ is T if the truth-value of $\psi(P)$ is T in at least one expansion of A to a structure for the language that includes every member of $\mathcal{L}$ and also P . It is F otherwise.

Once given the standard semantics a logical truth is defined to be a sentence true in every structure for the language of the sentence and a valid argument to be one such that there is no structure in which all the premises are true and the conclusion false, as for first-order logic. Moreover, since the notion of structure is unchanged from first-order logic, standard model-theoretic notions such as isomorphism and categoricity can be taken over unchanged from the model theory of first-order logic. (Two structures are said to be isomorphic if, roughly, there is a one-to-one correspondence between their universes that takes the interpretations of constant and relation

symbols in one to the corresponding interpretations in the other. A theory is said to be categorical if all its models are isomorphic.)

It is natural to turn now to a deductive system for second-order logic to characterize logical truths and valid arguments. This does not work out nearly so well for second-order logic as for first-order.

The core of any deduction system for second-order logic will be a first-order system supplemented with quantifier rules for the second-order quantifiers that exactly parallel those for the first-order ones, with n -placed relation symbols playing the role constant symbols did for first-order quantification in the case of second-order quantification over n -ary relations.

In standard second-order logic one assumes that for any formula ϕ with all of its free variables from the list $v_1, \dots, v_m, v_{m+1}, \dots, v_{m+n}$ and for any n individuals $b_1, \dots, b_n$, there exists an m -ary relation comprehending exactly those m -tuples of individuals $a_1, \dots, a_m$ which are such that

$$\phi[a_1, \dots, a_m, b_1, \dots, b_n]$$

holds, where the square brackets indicate that v_1 is replaced everywhere it occurs free in ϕ by a new constant symbol interpreted by a_1 in an expanded structure, and so on. The relation is said to be defined by ϕ with parameters $b_1, \dots, b_n$. More precisely, one assumes that the following axiom schema of comprehension holds:

$$(\forall v_{m+1}) \dots (\forall v_{m+n}) (\exists X_1^m) \\ (\forall v_1) \dots (\forall v_m) (X_1^m v_1 \dots v_m \leftrightarrow \phi),$$

where m and n are natural numbers and ϕ is a formula with all of its free variables among $v_1, \dots, v_{m+n}$. If the variables $v_1, \dots, v_m$ do not occur in the formula ϕ , then the comprehension axiom entails that the empty m -ary relation is in the domain of the second-order variables when ϕ does not hold of $b_1, \dots, b_n$, and that the universal m -ary relation is in the domain of the second-order variables when it does hold.

It is enough for many purposes to consider only the predicative axiom schema of comprehension, which requires additionally that ϕ be a formula that does not contain any second-order quantifiers.

One might add other axioms to a deduction system for second-order logic. However, there is no usable system of axioms that yields all the logical truths as theorems: if a deduction system for second-order logic is *sound* (that is, if it cannot be used to derive sentences that are not logical truths) and the theorems are recursively enumerable (that is, there are effective procedures for listing the axioms and checking deductions - see **Computability theory**), then it is incomplete in the sense that there is a logical truth that cannot be deduced in it.

The impossibility of providing a complete deduction system for second-order logic follows from Gödel's incompleteness theorems (see **Gödel's theorems §§3-4**) and the fact that there is a second-order sentence that is a categorical axiomatization of arithmetic (that is, an axiomatization such that any model is isomorphic to a model with universe the set of natural numbers with, say, for definiteness, the relation 'strictly less than' and the addition and multiplication operations - in less technical terms, this is an axiomatization that characterizes the natural numbers completely for all mathematical purposes). (A sentence that is a categorical axiomatization of arithmetic is described and discussed in §3 below.) Suppose the sentence Φ is a categorical axiomatization of arithmetic. A truth of arithmetic is just a sentence true of the natural numbers. Thus, a sentence ϕ is a truth of arithmetic just if it is true in every model of Φ , that is, just if the argument from Φ to ϕ is valid. But that means that a complete deduction system for second-order logic would yield a decision procedure for the truths of arithmetic: ϕ is true if deducible from Φ and not true if $\neg\phi$ is deducible from Φ , and a systematic enumeration of all possible deductions from Φ will eventually decide which of these is the case. Gödel's incompleteness theorem entails that no such decision procedure is possible, and so no such deduction system is possible.

2 Alternative semantics

Instead of requiring, as in the standard semantics, that second-order variables range over all relations on a domain, one can be more explicit about their range by specifying as part of a structure the domains of the second-order variables in addition to that for the first-order variables. A Henkin structure for second-order logic is, by definition, an ordinary structure (with a universe and an interpretation function) supplemented by a third

component: a domain function $\mathcal{D}$ on the natural numbers that takes each natural number n to a set of n -ary relations on the universe of the structure. For each natural number n , the set $\mathcal{D}(n)$ must contain every n -ary relation that is the interpretation of an n -placed relation symbol in the language of the structure. (This condition is frequently omitted. But no use is ever made of the Henkin structures that do not satisfy it.)

Let A be a Henkin structure with language $\mathcal{L}$, and let P and $\psi(P)$ be as above. In Henkin semantics, the clauses of the definition of truth for the quantifiers will be

(1') If the sentence ϕ is of the form $(\forall X_n^m)\psi$ for some m and n , then the truth-value of ϕ is T if the truth-value of $\psi(P)$ is T in every expansion of A to a Henkin structure for the language that includes every member of $\mathcal{L}$ and also P and interprets P by a member of $\mathcal{D}(n)$. It is F otherwise.

(2') If the sentence ϕ is of the form $(\exists X_n^m)\psi$ for some m and n , then the truth-value of ϕ is T if the truth-value of $\psi(P)$ is T in at least one expansion of A to a Henkin structure for the language that includes every member of $\mathcal{L}$ and also P and interprets P by a member of $\mathcal{D}(n)$. It is F otherwise.

One can define a Henkin logical truth to be a sentence true in every Henkin structure for the language of the sentence and a Henkin valid argument to be one such that there is no Henkin structure in which all the premises are true and the conclusion false. If one were to restrict attention to *full* Henkin models - in which for each n the value of $\mathcal{D}(n)$ is the set of all n -ary relations on the domain - one would recover the logical truths and valid arguments of standard second-order logic. There are possibilities intermediate between considering all Henkin structures and considering only full ones: for example, that of restricting attention to Henkin structures that are models of the comprehension schema.

The second-order quantification of Henkin semantics resembles first-order quantification in that the universe of quantification is specified. That suggests going even further - to a first-order semantics for (syntactically) second-order logic. The domains of n -ary relations of Henkin structures are here replaced by perfectly arbitrary domains of objects - no longer required to be relations on the universe of individuals - and the atomic formulas of the form $X_n^m t_1 \dots t_m$, which are now just of the semantic form of a list of $m + 1$ first-order constants and variables, are interpreted using what are called 'predication relations' (which should really be called 'relation relations' but are not, for obvious reasons). For example, $X_n^m v_1 \dots v_m$ becomes $P_m X_n^m v_1 \dots v_m$, where P_m is the predication relation for m -ary relations, which may be read 'the relation X_n^m holds of $v_1, \dots, v_m$ '.

Once there are predication relations, there is no need to introduce any other relations, except as objects in the first-order domains - all the relations can be taken to be members of the universes of the second-order variables. It is therefore possible to assume that the only relation symbols in a first-order structure for second-order logic are those for the predication relations and equality. This is not absolutely essential, but it does simplify matters considerably. A *first-order structure for second-order logic* will be constituted, as usual, of a non-empty universe of individuals, a domain function $\mathcal{D}$ on the natural numbers that takes each number to a non-empty set, and an interpretation function. The interpretation function interprets constant symbols as members of the various universes, and it interprets each predication relation symbol P_n as a relation that can hold only between a member of $\mathcal{D}(n)$ and n members of the universe of individuals, that is, the interpretation of each P_n is a set of $n + 1$ -tuples with first members from $\mathcal{D}(n)$ and all other members from the universe of individuals.

A Henkin structure can be converted into a first-order structure for second-order logic as follows. The domain function $\mathcal{D}$ is simply the domain function of the Henkin structure. The predication relations are based on membership relations: the n th predication relation holds between $R \in \mathcal{D}(n)$ and n objects just if the n -tuple formed by the n objects is a member of R . Each n -placed relation symbol must be replaced by a constant symbol that has the same interpretation, now viewed as a member of $\mathcal{D}(n)$. The semantics for second-order logic with first-order structures is just a standard first-order semantics for a multi-sorted structure, that is, one with many universes. The variables associated with a given universe (in our example, variables X_n^m are associated with $\mathcal{D}(n)$) are only interpreted by (or replaced by constant symbols interpreted by) members of that universe. The device of using many universes instead of one involves less than meets the eye: it is possible using entirely elementary methods to replace structures with many universes with ordinary single-universe structures that are equivalent in a suitably strong sense.

One can make use of intensional first-order structures for second-order logic, but here we restrict our attention to

extensional ones, which are, by definition, those that are models of the following schema:

$$(\forall X_1^n)(\forall X_2^n)((\forall v_1) \dots (\forall v_n) (X_1^n v_1 \dots v_n \leftrightarrow X_2^n v_1 \dots v_n) \rightarrow X_1^n = X_2^n).$$

The schema says that relations that hold of the same n -tuples are identical.

In extensional first-order structures for second-order logic, the set of n -tuples of which a relation holds fully determines the relation, and so extensional first-order structures for second-order logic might as well just be Henkin structures: every such structure $\mathcal{A}$ gives rise to a Henkin structure by dropping the predication relations and, for each n , replacing each member R of $\mathcal{D}(n)$ by the set of all n -tuples $(a_1, \dots, a_n)$ such that $(R, a_1, \dots, a_n)$ is in the n th predication relation of the first-order structure. Constant symbols that stand for objects in $\mathcal{D}(n)$ must be replaced by n -placed relation symbols that are interpreted as the relations corresponding to the given objects. The resulting Henkin structure, converted to a first-order structure by the procedure described above, yields a first-order structure suitably similar to the original structure $\mathcal{A}$.

3 Metamathematics

Dedekind observed that a set is finite if and only if there is no one-to-one correspondence between the set and any of its proper subsets. Thus, the following formula, $\Psi(X_1^1)$, on the standard semantics, expresses that the predicate X_1^1 holds of finitely many things:

$$\begin{aligned} &(\forall X_1^2)((\forall v_1)(X_1^1 v_1 \rightarrow (\exists v_2)(X_1^1 v_2 \wedge X_1^2 v_1 v_2)) \\ &\quad \wedge (\forall v_1)(\forall v_2)(\forall v_3)(X_1^2 v_1 v_2 \wedge X_1^2 v_1 v_3 \rightarrow v_2 = v_3) \\ &\quad \wedge (\forall v_1)(\forall v_2)(\forall v_3)(X_1^2 v_1 v_3 \wedge X_1^2 v_2 v_3 \rightarrow v_1 = v_2)) \\ &\rightarrow (\forall v_2)(X_1^1 v_2 \rightarrow (\exists v_1)(X_1^1 v_1 \wedge X_1^2 v_1 v_2))). \end{aligned}$$

Using this formula one can categorically axiomatize the natural numbers as a discrete linear order with a least element and no greatest element ('discrete' means that every element that has a predecessor (successor) has an immediate predecessor (successor)) in which the set of elements less than any element is finite. One can then supplement this with standard first-order axiomatizations of, for example, addition and multiplication. It is straightforward to give a first-order axiomatization of the discrete linear orders with a least element and no greatest element. The only second-order part of the categorical axiomatization is the use of the notion of finitude. That is our main concern here.

Note that on the Henkin interpretation of second-order logic $\Psi(X_1^1)$ does not define finitude: in order for the definition to work, it is crucial that the domain of the variable X_1^2 include relations that yield *all* one-to-one correspondences from X_1^1 to itself. To give a trivial example, $\Psi(X_1^1)$ holds vacuously of any set - including an infinite one - in a Henkin structure in which there simply are not any binary relations that relate objects in the set to other objects in the set. It is clear that this example may in some sense be taken to be artificial. Indeed, the constraints we have placed on which Henkin structures are to be allowed exclude ones in which there are no binary relations whatever on some part of the universe. But if constraints are to be imposed axiomatically - that is, by restricting attention to Henkin models of some set of axioms - then no constraints help to ensure that Ψ , or indeed any other formula, characterizes finitude. Any suitable axiomatic constraints would, as was in effect shown in §2, carry over into first-order axiomatic constraints such that finitude could be characterized over the first-order structures for second-order logic that satisfy them.

But the notion of finitude cannot be characterized axiomatically in first-order logic, since first-order logic is *compact*. The compactness theorem for first-order logic states that if every finite subset of a set of sentences of first-order logic has a model, then so does the whole set. Suppose a set S of sentences of first-order logic has models in which the interpretation of a predicate symbol P is of arbitrarily large finite size. Let S' be the set of sentences including those in S plus sentences of first-order logic formalizing 'There is at least one thing v such that $P(v)$ ', 'There are at least two things v such that $P(v)$ ', and so on. Then every finite subset of S' has a model. But then by the compactness theorem the whole set has a model, in which the extension of P must be infinite. So if a set S of sentences of first-order logic has models in which the interpretation of P is of every finite size then it has a model in which the interpretation of P is infinite. It fails to characterize finitude. (See [Model theory](#).) Thus, the

notion of finitude cannot be axiomatically characterized in second-order logic interpreted using either first-order or Henkin structures.

By a contraposition of this argument, since finitude can be characterized in standard second-order logic, standard second-order logic is not compact. Thus, a sentence of standard second-order logic can be a logical consequence of infinitely many sentences of standard second-order logic without being a consequence of any finite subset of them. This is a very strong failure of the very idea of having a deduction system for standard second-order logic.

In §1, it was shown that no deduction system captured the logical truths of standard second-order logic. It now follows that standard second-order logic is also difficult to characterize in a second sense: the Henkin structures that are associated with standard second-order logic - the full Henkin structures - cannot be characterized axiomatically, in the sense that there is no set of axioms, not even the set of all the logical truths of standard second-order logic, that hold of all and only the full Henkin structures. That follows because if there were such a set of axioms (since finitude can be characterized in standard second-order logic) they could be used to characterize axiomatically the notion of finitude.

The situation is entirely similar with many other familiar mathematical notions: like finitude, many notions that mathematicians seem to feel they understand have adequate definitions in standard second-order logic but not in first-order logic.

See also: [Logical and mathematical terms, glossary of](#); [Second-order logic, philosophical issues in](#)

SHAUGHAN LAVINE

References and further reading

- Church, A.** (1956) *Introduction to Mathematical Logic*, Princeton, NJ: Princeton University Press. (An influential introduction to second- (and first-) order logic.)
- Ebbinghaus, H.-D., Flum, J. and Thomas, W.** (1984) *Mathematical Logic*, Berlin and New York: Springer, 2nd edn, 1994. (The mathematical results about second-order logic found in this article are proved in detail. Results characterizing first-order logic - and hence distinguishing it from second-order and other logics - are presented in an unusually accessible manner.)
- Field, H.** (1994) 'Are our Logical and Mathematical Concepts Highly Indeterminate?', in P.A. French, T.E. Uehling, Jr and H.K. Wettstein (eds) *Midwest Studies in Philosophy*, vol. 19, *Philosophical Naturalism*, Notre Dame, IN: University of Notre Dame Press, 391-429. (The ways in which second-order logic does and does not help in the determination of our logical and mathematical concepts are analysed in some detail, and a number of arguments and positions are carefully laid out.)
- Hilbert, D. and Ackermann, W.** (1928) *Grundzüge der theoretischen Logik*, Berlin: Springer, 2nd edn, 1938; trans. L.M. Hammond, G.G. Leckie and F. Steinhardt, *Principles of Mathematical Logic*, ed. R.E. Luce, New York: Chelsea, 1950. (A classic introduction to logic, including second-order logic; quite clear and detailed.)
- Lavine, S.** (1994) *Understanding the Infinite*, Cambridge, MA: Harvard University Press. (Though the main focus of this work is the theory of sets, it includes a detailed discussion of how a part of second-order logic plays a role in fixing mathematical concepts.)
- Monk, J.D.** (1976) *Mathematical Logic*, Graduate Texts in Mathematics 37, New York: Springer. (This text includes a concise but readable introduction to the basics of the syntax and semantics of second- and higher-order logic.)
- Moore, G.H.** (1988) 'The Emergence of First-Order Logic', in W. Aspray and P. Kitcher (eds) *History and Philosophy of Modern Mathematics*, Minneapolis, MN: University of Minnesota Press, 95-135. (A stellar synoptic treatment of the emergence of first-order logic - and hence, as a sort of residue, second-order logic - from earlier systems of logic that classified things rather differently. This article provides an excellent entry point into the modern history of the subject, with a very useful bibliography.)
- Shapiro, S.** (1991) *Foundations without Foundationalism*, Oxford: Clarendon Press. (A comprehensive work on basics of second-order logic and the ways in which it is relied on in much of ordinary mathematical practice. Extensive bibliography.)

Second-order logic, philosophical issues in

Typically, a formal language has variables that range over a collection of objects, or domain of discourse. A language is 'second-order' if it has, in addition, variables that range over sets, functions, properties or relations on the domain of discourse. A language is third-order if it has variables ranging over sets of sets, or functions on relations, and so on. A language is higher-order if it is at least second-order.

Second-order languages enjoy a greater expressive power than first-order languages. For example, a set S of sentences is said to be categorical if any two models satisfying S are isomorphic, that is, have the same structure. There are second-order, categorical characterizations of important mathematical structures, including the natural numbers, the real numbers and Euclidean space. It is a consequence of the Löwenheim-Skolem theorems that there is no first-order categorical characterization of any infinite structure. There are also a number of central mathematical notions, such as finitude, countability, minimal closure and well-foundedness, which can be characterized with formulas of second-order languages, but cannot be characterized in first-order languages.

Some philosophers argue that second-order logic is not logic. Properties and relations are too obscure for rigorous foundational study, while sets and functions are in the purview of mathematics, not logic; logic should not have an ontology of its own. Other writers disqualify second-order logic because its consequence relation is not effective - there is no recursively enumerable, sound and complete deductive system for second-order logic.

The deeper issues underlying the dispute concern the goals and purposes of logical theory. If a logic is to be a calculus, an effective canon of inference, then second-order logic is beyond the pale. If, on the other hand, one aims to codify a standard to which correct reasoning must adhere, and to characterize the descriptive and communicative abilities of informal mathematical practice, then perhaps there is room for second-order logic.

1 First-order and second-order logic compared

The downward Löwenheim-Skolem theorem states that if S is a set of first-order sentences that has an infinite model, then S has a model whose domain is at most countably infinite (or the cardinality of S , whichever is larger). The upward Löwenheim-Skolem theorem states that if S is a set of first-order sentences such that, for each natural number n , S has a model whose domain has at least n elements, then for every infinite cardinal κ , S has a model whose domain has cardinality at least κ . It follows that no set of first-order sentences that has an infinite model is categorical. (A set of sentences is said to be categorical if any two models satisfying it are isomorphic.)

There is a complete, sound and effective deductive system D for first-order logic (see [Predicate calculus](#)). That is, if S is a set of sentences and ϕ is a single sentence, then ϕ is deducible from S in D if and only if ϕ holds in every structure that satisfies every member of S . Finally, first-order logic is compact, in that for every set S of first-order sentences, if every finite subset of S has a model, then S itself has a model.

(In standard semantics for second-order languages, the set or predicate variables range over the entire power set of the domain. It follows that the domain of discourse determines the range of the other variables. There is an alternative semantics, due to Leon Henkin (1950), in which this does not hold. Each model consists of not only a domain of discourse, but also separate ranges for the predicate, relation and function variables. With Henkin semantics, second-order logic is complete and compact, and the Löwenheim-Skolem theorems hold. In effect, with Henkin semantics, second-order languages are treated as multi-sorted first-order languages. The foregoing discussion presupposes standard semantics.)

Second-order logic has none of the above properties. The upward Löwenheim-Skolem theorem fails because there is a categorical characterization AR of the natural numbers. Every model of AR is countably infinite. The downward Löwenheim-Skolem theorem fails because there is a second-order categorical characterization AN of the real numbers. Every model of AN has the cardinality of the continuum, and so is uncountable (see [Cantor's theorem](#); [Continuum hypothesis](#)). A corollary of these features, and Gödel's incompleteness theorem, is that second-order logic is not complete (see [Gödel's theorems §1](#)). There is no effective, sound and complete deductive system for second-order logic. Likewise, second-order logic is not compact.

The sentence AR characterizing the natural numbers is a straightforward formalization of common informal

axiomatizations of arithmetic. The ‘second-order’ item is the induction principle, stating that if a property (set) P of numbers holds for (contains) zero and is closed under the successor function, then P holds for (contains) all natural numbers. In symbols,

$$\forall P[(P0 \ \& \ \forall x(Px \rightarrow Psx)) \rightarrow \forall xPx].$$

In first-order arithmetic, this principle is replaced by a scheme. Every instance of

$$[\phi(0) \ \& \ \forall x(\phi(x) \rightarrow \phi(sx))] \rightarrow \forall x\phi(x)$$

in which ϕ is a formula in the language of first-order arithmetic, is an axiom. Thus, first-order arithmetic has infinitely many axioms.

In the characterization *AN* of real analysis, the second-order item is the principle of completeness, stating that every non-empty set of real numbers with an upper bound has a least upper bound:

$$\begin{aligned} (\forall P)\{(\exists xPx \ \& \ \exists x\forall y(Py \rightarrow y \leq x)) \\ \rightarrow \exists x[\forall y(Py \rightarrow y \leq x) \ \& \\ \forall z(\forall y(Py \rightarrow y \leq z) \rightarrow x \leq z)]\} \end{aligned}$$

First-order real analysis is obtained by replacing the completeness axiom with the completeness scheme,

$$\begin{aligned} (\exists x\phi(x) \ \& \ \exists x\forall y(\phi(y) \rightarrow y \leq x)) \\ \rightarrow \exists x[\forall y(\phi(y) \rightarrow y \leq x) \ \& \\ \forall z(\forall y(\phi(y) \rightarrow y \leq z) \rightarrow x \leq z)] \end{aligned}$$

one instance for each formula ϕ that contains neither x nor z free.

The difference between, say, second-order real analysis and its first-order counterpart is that in the former one cannot directly state that every non-empty bounded set has a least upper bound. Instead, there is a separate principle for each such set which is definable by a formula in the language of first-order analysis. Since there are sets which are not definable, the first-order theory has models which are not isomorphic to the real numbers. Such structures are called ‘nonstandard models’ (see [Löwenheim-Skolem theorems and nonstandard models](#)).

One common characterization of finitude, due to Richard Dedekind, is that a set is finite if there is no one-to-one function from the set to a proper subset of itself. This can be rendered as a second-order sentence with no non-logical terminology:

$$\begin{aligned} \text{FIN}(X): \sim\exists f[\forall x(Xx \rightarrow Xfx) \ \& \\ \forall x\forall y((Xx \ \& \ Xy \ \& \ fx = fy) \rightarrow x = y) \\ \ \& \ \exists z(Xz \ \& \ \forall x(Xx \rightarrow fx \neq z))] \end{aligned}$$

It follows from the compactness property that there is no first-order characterization of finitude.

2 Defence

It is widely held, often implicitly, that mathematicians succeed in describing various structures, up to isomorphism, and in communicating information about them. When various mathematicians refer to ‘the natural numbers’, ‘the real numbers’, ‘Euclidean space’ and so on, they are talking about the same structures. Similarly, there is little doubt that such mathematical notions as finitude and well-foundedness are clear, definite and unequivocal. In short, it is agreed that the *informal* language of mathematics is sufficient for ordinary description and communication of the basic structures and notions.

One purpose of logic, model theory in particular, is to ‘model’ the semantics of the informal languages of mathematics. Thus, one desideratum is that logic should somehow register the successful description and communication of mathematical structures and notions. As we have seen, first-order logic falls short of this. As Barwise and Feferman put it, ‘As logicians, we do our subject a disservice by convincing others that logic is first-order and then convincing them that almost none of the concepts of modern mathematics can really be captured in first-order logic’ (1985: 5).

A further claim is that it is unnatural to rely on infinitely many axioms, when a single second-order axiom is available. For example, if someone is asked why they believe that the second-order completeness axiom holds for the real numbers, the reply might be that the principle characterizes the numbers, perhaps citing the categoricity result. However, if asked why they believe every instance of the first-order completeness scheme, the answer is not that simple. One cannot claim that the scheme characterizes the real numbers since, as we have seen, the scheme does not characterize the numbers. Georg Kreisel suggests that the reason that mathematicians believe the instances of the various schemes is that they are consequences of the respective second-order axiom (1967). Indeed, in common deductive systems for second-order logic, each instance of each scheme can be deduced from the corresponding axiom. An opponent might retort that one can see from the scheme itself that all of its instances are true. This might be called a ‘metalinguistic’ move, since it invokes the scheme itself at the level of justification. Notice that this manoeuvre involves a generalization, since it refers to *all* instances of the scheme. It remains to be seen whether this generalization is any more problematic than the generalization over predicates, properties or sets. Notice also that the scheme itself is an infinite structure, and cannot be characterized in a first-order metalanguage.

Another problem with the schemes is that each one is tied to the ingredients of the particular language in use at the time. For example, if the induction scheme is formulated in a language whose only non-logical term is the successor symbol, then a very weak, decidable theory results. To obtain a richer system, closer to classical arithmetic, symbols for addition and multiplication must be added, and the induction scheme must itself be expanded to include instances with the new terminology. On the other hand, the second-order induction axiom, which contains only the successor symbol, is sufficient to characterize the natural numbers. Moreover, addition and multiplication can be defined in second-order arithmetic, as can any recursive function. Beyond that, new terminology can be added as needed, without having to modify the very definition of the natural numbers. This fits the intuition that one cannot tell in advance what resources, and thus what terminology, are needed to shed light on the natural numbers, even after the structure has been characterized.

In short, ‘the natural numbers’ has an informal interpretation that outruns what is captured in first-order arithmetic, and this informal interpretation plays a central role in actual mathematical practice. Alonzo Church agrees:

our definition of the consequences of a system of postulates...can be seen to be not essentially different from [that] required for the...treatment of classical mathematics.... It is true that the non-effective notion of consequence, as we have introduced it...presupposes a certain absolute notion of ALL propositional functions of individuals. But this is presupposed also in classical mathematics, especially classical analysis

(1956: 326)

3 Prosecution

A number of authors have either rejected second-order logic outright, or else claimed that it is not logic, but rather an obscure form of mathematics. One broad attack is aimed at ‘traditional’ versions of second-order and higher-order systems that have variables ranging over intensional items such as properties, attributes and propositional functions (see **Intensional entities**). There is no consensus on which properties, attributes and so on exist, and there is no consensus on when two properties are identical or distinct. These sticky metaphysical matters should not interfere with serious work in the foundations of mathematics, the argument goes. Mathematics itself is reasonably clear, and we should not invoke obscurity in order to develop foundations for it.

This argument does not affect contemporary systems. The items in the range of the second-order variables in standard semantics are extensional - they are sets or classes. For any collection *C* of elements of the domain of discourse, there is an item in the range of the ‘property’ variables whose elements are the members of *C*. The development of modern set theory shows that the extensional notion of class is at least relatively clear (see **Set theory**).

The problem now, however, is that by invoking set-theoretic notions in the model theory, we may have crossed a border from logic into mathematics. W.V. Quine calls second-order logic ‘set theory in disguise’ and ‘set theory in sheep’s clothing’. The metaphor suggests that, unlike grizzly mathematics, logic is, or ought to be, a mere ‘sheep’.

Quine and others have brought a number of arguments for this judgment. One involves ontological commitment

(see **Ontological commitment**). A first-order theory, with an intended interpretation, has variables ranging over a given domain of discourse. A second-order theory, with standard semantics, has variables ranging over the entire power set of the domain, a larger collection. Moreover, the usual metatheory for second-order languages, where we turn for information about the power set operation, uses Zermelo-Fraenkel set theory (see **Set theory, different systems of §3**), which does have a vast ontology. As Quine puts it, ‘Set theory’s staggering existential assumptions are...hidden...in the tacit shift from schematic predicate letter to quantifiable set variable’ ([1970] 1986: 68).

Because of the expressive resources of second-order languages, including the terminology for sets, there are a number of rather substantial mathematical statements that can be formulated in second-order languages. There is, for example, a sentence that is a logical truth if and only if the generalized continuum hypothesis is true, and there is a sentence that is a logical truth if and only if the axiom of choice holds. But it is on the face of it counterintuitive to hold that these statements are logically true or logically false.

It follows from Gödel’s completeness theorem that the consequence relation of first-order logic is effective. The set of first-order logical truths, for example, is recursively enumerable. On the other hand, the set of second-order logical truths is not. The set is not even definable in (second-order) arithmetic, although it is definable in set theory. Again, we seem to be beyond the purview of logic (see **Non-constructive rules of inference**).

In summary, many of those who hold that second-order logic is not logic seem to agree that the *informal* language of mathematics is somehow adequate to describe and communicate the various structures and notions, but they add that logic must fail where informal mathematics succeeds. There is another line of argument, less popular today, that rejects the presupposition that the mathematical structures and notions in question are unequivocal. Accordingly, the Löwenheim-Skolem theorems indicate that there is no unambiguous notion of ‘finite’, ‘countable’, ‘natural number’ and so on. All of these notions are relative to a background (first-order) set theory. This sceptical view is sometimes called ‘Skolemite relativism’. Advocates of it adopt first-order logic because they claim that its model theory accurately reflects the ontological/epistemic/semantic situation. There is nothing unequivocal to capture or describe with either formal or informal languages of mathematics.

4 Settlement?

By itself, a border dispute concerning mathematics may not be all that interesting. The deeper matters here concern the role and purposes of logical theory. Why do we develop logical systems in the first place? Historically, a number of goals have been articulated. One of these is to present a calculus, a deductive system to codify or serve as a canon of correct reasoning. In light of incompleteness, this goal is not served by the consequence relation of second-order logic (as long as a system must be effective, or recursive, in order to qualify as a calculus; see **Church’s thesis**). There are, to be sure, good deductive systems for second-order languages, but they are all incomplete for standard semantics.

On the other hand, it is widely held that deductive systems must themselves adhere to a prior notion of logical consequence. This notion is often understood in modal terms. Near the beginning of the *Prior Analytics*, Aristotle defines a ‘syllogism’ to be ‘a discourse in which, certain things having been supposed, something different from the things supposed results *of necessity*’ (see **Logic, ancient §§1-3**). In advance of reflection, there is no reason to expect that the pre-theoretic, modal notion of logical consequence is effective, and that there is a complete deductive system for it. Advocates of second-order logic claim that the second-order consequence relation captures an underlying notion of logical consequence, one to which deductive systems must answer.

There is a trend in philosophy to understand at least some modality in terms of language. In particular, it is held by many that necessities are grounded in the meanings of terms. In this spirit, logical consequence is sometimes defined in terms of the meanings of a certain collection of terms, the so-called ‘logical terminology’. This is consonant with the slogan that logical consequence is a matter of ‘form’. The logical terminology of a discourse is what marks its form, while the non-logical items determine content. From this perspective, the issue of second-order logic is whether the membership (or predication) relation and bound variables ranging over classes are logical. This, of course, is another border dispute. It should be settled by examining the various systems, to determine which are the most useful and insightful accounts of logical consequence. Perhaps one can be eclectic (see **Logical constants; Logical form**).

Second-order, and even higher-order, languages have found applications outside of logic and the foundations of mathematics. Two noteworthy examples are Montague grammar, for natural language (1974; see [Montague, R.M.](#)), and Aldo Bressan's higher-order modal logic for the language of science (1972). Although, as noted, many logicians have come to see first-order logic as too weak, it is still widely held that the consequence relation of second-order logic is too intractable for useful mathematical study. A growing research programme in mathematical logic is to develop and study alternatives which are, in a sense, intermediate between first-order and second-order.

See also: [Consequence, conceptions of](#); [Logical and mathematical terms, glossary of](#); [Second- and higher-order logics](#); [Theory of types](#)

STEWART SHAPIRO

References and further reading

- Barwise, J. and Feferman, S.** (eds) (1985) *Model-Theoretic Logics*, New York: Springer. (Extensive treatment of many logics, some of which are intermediate between first-order and second-order. Quite technical, for the most part. Extensive bibliography.)
- Bressan, A.** (1972) *A General Interpreted Modal Calculus*, New Haven, CT: Yale University Press. (An extensive, higher-order modal logic, put to a number of philosophical uses.)
- Church, A.** (1956) *Introduction to Mathematical Logic*, Princeton, NJ: Princeton University Press. (Still an excellent introduction to both first-order and second-order logic.)
- Heijenoort, J. van** (ed.) (1967) *From Frege to Gödel*, Cambridge, MA: Harvard University Press. (Collection of many original papers in logic, translated into English. Includes papers by Frege, Peano, Dedekind, Cantor, Hilbert, Russell, Zermelo, Löwenheim, Skolem, Weyl, Ackermann, von Neumann, Bernays and Gödel.)
- Henkin, L.** (1950) 'Completeness in the Theory of Types', *Journal of Symbolic Logic* 15: 81-91. (Henkin's alternative semantics, which treats second-order languages as multi-sorted first-order languages, with separate ranges for the predicate, relation and function variables.)
- Hilbert, D. and Ackermann, W.** (1928) *Grundzüge der theoretischen Logik*, Berlin: Springer; trans. L. Hammond, G. Leckie and F. Steinhardt, *Principles of Mathematical Logic*, New York: Chelsea, 1950. (Historical source; quite readable.)
- Kreisel, G.** (1967) 'Informal Rigour and Completeness Proofs', in I. Lakatos (ed.) *Problems in the Philosophy of Mathematics*, Amsterdam: North Holland, 138-86. (Among other things, a multi-faceted philosophical defence of the use of higher-order languages in foundational studies. Moderate difficulty.)
- Montague, R.** (1974) *Formal Philosophy*, ed. R. Thomason, New Haven, CT: Yale University Press. (Higher-order logic, put to linguistic use. The start of an extensive research programme.)
- Quine, W.V.** (1970) *Philosophy of Logic*, Englewood Cliffs, NJ: Prentice Hall; 2nd edn, Cambridge, MA: Harvard University Press, 1986, esp. ch. 5. (Argument against second-order logic. Lucid.)
- Shapiro, S.** (1991) *Foundations Without Foundationalism: A Case for Second-Order Logic*, Oxford: Oxford University Press. (Extensive development and justification for higher-order logic. Philosophical, mathematical and historical treatment. Most of the book requires some background in logic. Extensive bibliography.)

Semantic paradoxes and theories of truth

The Cretan philosopher Epimenides said that Cretans always lie. Assuming, for the sake of argument, the mendacity of all other statements by Cretans, we get a paradox: if what Epimenides said was true, it must have been a lie, whereas if what he said was a lie, it would have made his statement true. The citizens of Crete have long since forgiven the insult, but semantics has never recovered.

Alfred Tarski perceived the consequences of Epimenides' paradox with particular clarity. Our common-sense intuitions about truth follow the paradigm: 'Snow is white' is true if and only if snow is white. As Tarski rigorously shows, if the language we are describing (the object language) is the same as the language in which we are formulating our theory (the metalanguage), this paradigm will be inconsistent with the rudimentary laws of syntax. The conclusion Tarski drew was that, if we are to develop a satisfactory theory of truth, our metalanguage must be essentially richer in expressive power than the object language. Since there is no human language essentially richer than English (or any other natural language), there can be no satisfactory theory of truth for English.

One earnestly hopes that this is not the end of the matter. Tarski's analysis leaves open the prospect that we can develop a fully satisfactory theory of truth for a substantial fragment of English; also the prospect that we can develop a theory of truth for English as a whole which, while not fully satisfying our intuitions, is none the less useful and illuminating.

Both prospects have been substantially advanced by Saul Kripke's 'Outline of a Theory of Truth', which exploits the idea that there are truth-value gaps, statements which are neither true nor false, and that Epimenides' insult was one of them.

Invocation of truth-value gaps does not resolve the paradox in any straightforward way. If we let the phrase 'the simple liar sentence' refer to the sentence 'The simple liar sentence is false', we see that we can readily account for the paradoxical features of the sentence by declaring the sentence neither true nor false; but if we let the strengthened liar sentence be 'The strengthened liar sentence is not true', we get a sentence we cannot dispose of so tidily. If the strengthened liar is neither true nor false, then it is not true; but that it is not true is precisely what the sentence says.

Truth-value gaps have not vanquished the liar paradox. Nor have any of the alternatives, the most prominent of which are a contextualist account, which sees the English word 'true' as radically ambiguous, and so-called 'revision theory', which investigates the cyclic reasoning that occurs when we try to evaluate the simple liar sentence: if the sentence is true, then it must be false; but if, then, it is false, it must be true; and so on. While these approaches have not eliminated the paradox, they have opened new approaches that have greatly improved our prospects for finding a comfortable way to live with it.

1 Tarski's requirement of a richer metalanguage

Kurt Gödel (1931) proved the incompleteness of a version of Peano Arithmetic by associating a code number, denoted $\ulcorner \phi \urcorner$, with each sentence ϕ of the language of arithmetic, then constructing a sentence χ which denied that $\ulcorner \chi \urcorner$ was the code of a theorem (see Gödel's theorems). Tarski (1935) used Gödel's construction to give a formalized version of the liar paradox. If $Tr(x)$ were a formula of the language of arithmetic which represented the set of code numbers of true sentences, we would expect to have

$$Tr(\ulcorner \phi \urcorner) \leftrightarrow \phi(T)$$

for each sentence ϕ . But Gödel's methods can be used to construct a sentence λ such that $\lambda \leftrightarrow \neg Tr(\ulcorner \lambda \urcorner)$ is a theorem. We conclude that there is no formula of the language of arithmetic which represents the set of code numbers of true sentences of the language of arithmetic. Indeed, this result holds not only for the language of arithmetic but for any language which includes the language of arithmetic.

Arithmetic plays an inessential role here. Quine (1946) shows that we can encode the language of arithmetic into any language which can provide a rudimentary theory of syntax. We are forced to conclude that, if $\mathcal{L}$ is a language that is able to describe its own syntax, then there is no open sentence $Tr(x)$ of $\mathcal{L}$ such that the schema (T) is consistent with the laws of syntax; here $\ulcorner \phi \urcorner$ is a standard name of ϕ which describes the sentence in terms of its

syntactic structure. On the other hand, any semantic theory which adequately represents our intuitive understanding of truth must surely imply all the (T)-sentences. We conclude that there is no possibility, consistent with the laws of syntax, of finding an open sentence of $\mathcal{L}$ which adequately represents the set of truths of $\mathcal{L}$. The conclusion Tarski drew, and which nearly everyone has accepted, is that, to develop a satisfactory theory of truth for $\mathcal{L}$ we must work within a metalanguage richer than $\mathcal{L}$ in expressive power.

2 *Principia Mathematica* and its successors

Whitehead and Russell responded to Epimenides' paradox by proposing a layered multiplicity of notions of truth. Simple statements of fact enjoy the first order of truth; attributions of truth at order one will be true or false at order two; and so on. An attribution of n th order truth must always have order higher than n . In this way (though they did not give details), Whitehead and Russell were able to develop the semantics of *Principia Mathematica* without contradictions. *Principia Mathematica* was designed to resolve a variety of paradoxes in addition to the liar. Here we shall follow contemporary practice in separating the semantic from the set-theoretic paradoxes (as recommended by Ramsey 1925), and in focusing on the liar as our paradigmatic representative of the former.

Tarski's approach to the liar was similar to Whitehead and Russell's, but simplified from the intricacies of the theory of types. Tarski spoke of a hierarchy of languages. The ground language has no semantic predicates, while the $(n + 1)$ th language is derived from the n th language by adding a predicate true_n whose extension is the true sentences of the n th language.

Whitehead and Russell proposed to apply their analysis to natural languages by treating the English word 'true' as ambiguous; contradictions arise from failing to recognize the ambiguity.

Their proposal has been vigorously pursued, by Tyler Burge (1979), Charles Parsons (1974), Jon Barwise and John Etchemendy (1987) and others, who have treated 'true' as an indexical term, like 'here' and 'now', whose extension is determined in part by the context of utterance. If everything Plato says is either true_{12} or false_{12} , the use of 'true' in 'Everything Plato says is true' will be contextually determined to mean ' true_{13} '. One quickly requires transfinite indices, going beyond the finite levels of *Principia Mathematica*.

Paradoxes arise when the customary rules for placing subscripts falter. Ordinarily, the subscript taken by 'false' in the sentence 'Everything a Cretan says is false' will be greater than the subscripts supplied to any occurrence of 'true' or 'false' in any utterance by a Cretan. But it is not possible to observe this rule when the sentence is uttered by a Cretan. If we take the word 'false' in Epimenides' statement 'Everything a Cretan says is false' to mean ' false_α ', we find that his statement is neither true_α nor false_α , since the predicates ' true_α ' and ' false_α ' only apply to sentences with subscripts less than α .

From a higher level of reflection of which Epimenides himself was in principle incapable, we can see that, since Epimenides' statement that everything a Cretan says is false_α is not false_α , we can make out a sense in which the sentence 'Everything a Cretan says is false' is false, namely, 'Everything a Cretan says is false_α ' is $\text{false}_{\alpha+1}$. We should not say, however, that what Epimenides said was false after all, rather, the sentence Epimenides uttered expresses something false when we say it, though in Epimenides' mouth it was neither true nor false.

Talk of levels of reflection is something the theorist engages in when describing the language, but not something speakers of the language engage in. If this were not so, Epimenides could say 'Nothing a Cretan says is true at any level of reflection' and reinstate the paradox. Thus we still submit to Tarski's requirement that the metalanguage have expressive powers not available in the object language. Again, the theorist can make the subscripts under 'true' and 'false' explicit, and use variables as subscripts. The speakers of the language cannot do this, otherwise they could get a contradiction by formulating the explicit liar ('The explicit liar is not true_α , for any α '). Contextualists have not overturned Tarski's requirement that the semantics of a language be developed in a richer metalanguage, nor, for the most part, have they attempted to do so. Instead they have tried to show that, contrary to first impressions, Tarski's requirement is relaxed enough to make room for virtually everything speakers of English want to say.

3 Kripke's theory of truth

We have been pretending that, apart from one notorious remark by Epimenides, everything ever said by a Cretan is

outright false. This pretence is, of course, contrary to fact, and, once we drop the pretence, Epimenides' statement is no longer problematic. If we hear a Cretan remark that snow is white, we can conclude at once that not all Cretans always lie, hence Epimenides' statement is false. To reach this judgment, it is not necessary to examine everything a Cretan has ever said; it is enough to find one Cretan truth.

The assignment of levels we get from *Principia Mathematica* is unnecessarily extravagant. It is based upon the idea that, to determine the truth of 'Everything a Cretan says is false', we need to take account of everything any Cretan ever says, so it tries, unsuccessfully, to assign to the sentence a level higher than the level assigned to any Cretan statement. But to determine that Epimenides' statement is false, we do not need to take account of everything a Cretan ever says, we only need to find one truth uttered by a Cretan. So, if some Cretan statement is true at the first level, we can declare Epimenides' statement false at the second level, without waiting to examine a raft of irrelevant further statements.

This adjustment in the assignment of levels effects a remarkable simplification and clarification of the hierarchical structure. In describing the construction, we shall utilize a first-order language with primitive connectives 'and', 'not' and 'all', other connectives defined as usual. All terms other than 'Tr' are assumed already to have a fixed meaning. We assume that our universe of discourse is a set, and we suppose that every member of the universe is named by some constant, expanding the language if necessary to ensure this. (Alternatively, we could define truth in terms of satisfaction, as Tarski does; but this would complicate matters.) We assume that the language can describe its own syntax, either directly or via a Gödel coding.

A non-semantic atomic sentence will be either true_0 or false_0 .

A sentence of the form $\text{Tr}(c)$ is $\text{true}_{\alpha+1}$ if c denotes a sentence which is true_α , and $\text{false}_{\alpha+1}$ if c denotes a sentence which is false_α . If the individual denoted by c is not a sentence, $\text{Tr}(c)$ is false_0 .

If ϕ is true_α and ψ is true_β , the conjunction $(\phi \wedge \psi)$ will be $\text{true}_{\max(\alpha, \beta)}$. If either conjunct is false_γ for some γ , the conjunction is false_γ for the least such γ .

A negation is true_α if the negatum is false_α and false_α if the negatum is true_α .

A universally quantified sentence is true_α if α is the least ordinal such that every substitution instance of the sentence is true_β for some $\beta \leq \alpha$. If some substitution instance of $\psi(x)$ is false_γ , then $(\forall x)\psi(x)$ is false_γ for the least such γ .

Kripke's proposal (1975) is that a sentence be regarded as true if it is true_α for some α , false if it is false_α for some α , and undetermined otherwise. Applying his construction to a suitably regimented fragment of English, we find that it manages narrowly to skirt the edge of paradox without falling into contradiction, and the truth-values it supplies, when it supplies truth-values, are invariably the intuitively correct ones. For example, suppose that Socrates says only

Everything Plato says is true.

while Plato says

Not everything Socrates says is true.

Everything Epimenides says is true.

and Epimenides says

Nothing a Cretan says is true.

Snow is white.

Epimenides' second statement is true_0 , so his first statement is false_1 . Hence Plato's second statement is false_2 , Socrates' statement is false_3 , and Plato's first statement is true_4 .

Not only does Kripke's theory get the right truth-values, but, more importantly, it explains them. Sentences declared true are true because they are grounded in the non-semantic facts. Thus a sentence is true_α for some α just in case it is derivable from true atomic and negated atomic non-semantic sentences by the following rules,

while a sentence is false_α , for some α , just in case its negation is derivable:

‘ Tr ’-introduction: If c denotes a sentence ϕ , infer $Tr(c)$ from ϕ .

‘ $\neg Tr$ ’-introduction: If c denotes a sentence ϕ , infer $\neg Tr(c)$ from $\neg\phi$. Also, if c denotes a non-sentence, you may write $\neg Tr(c)$.

Conjunction introduction: Infer a conjunction from its conjuncts.

Negated conjunction introduction: Infer the negation of a conjunction from the negation of either conjunct.

Double negation introduction: From a sentence, infer the negation of its negation.

Universal quantification introduction: Infer a universal quantified statement from all of its instances. (This is an infinitary rule.)

Negated universal quantification introduction: Infer the negation of a universally quantified statement from the negation of any of its instances.

4 The fixed-point theorem

We have been speaking of ‘true’ as an ambiguous predicate. It is perhaps simpler, and certainly closer to Kripke’s original way of thinking, instead to regard ‘true’ as a univocal predicate in a language with a nonclassical logic. We employ Kleene’s ‘strong three-valued logic’ (see Kleene 1952: §54), according to which a predicate has an ‘extension’, consisting of those things to which it is determined that the predicate applies, and an ‘anti-extension’, consisting of those things to which it is determined that the predicate does not apply. The two cannot overlap, but there may be things to which it remains unsettled whether the predicate applies. An atomic sentence $Rc_1 \dots c_n$ is true if the n -tuple consisting of the things denoted by c_1 to c_n lies in the extension of R , false if it lies in the anti-extension, and truth-valueless otherwise. A conjunction is true if both conjuncts are true and false if either conjunct is false. A negation is true if the negatum is false and false if the negatum is true. A universally quantified statement is true if every instance is true and false if at least one instance is false.

In the application we are interested in, all the terms other than ‘ Tr ’ already have a fixed interpretation in which the anti-extension is the complement of the extension, so all we need to determine is the extension and the anti-extension of ‘ Tr ’. Where A and B are non-overlapping sets, let $\Gamma(A, B)$ be the pair $\langle C, D \rangle$, where C is the set of sentences which are declared true if A is assigned as the extension of ‘ Tr ’ and B as its anti-extension; and D is the union of the set of non-sentences and the set of sentences declared false when A is assigned as the extension of ‘ Tr ’ and B as its anti-extension. $\langle A, B \rangle$ is a ‘fixed point’ if and only if $\langle A, B \rangle = \Gamma(A, B)$.

Kripke’s construction can now be succinctly characterized. Let C be the set of sentences which are true_α for some α , and let D be the set consisting of the non-sentences and the sentences which are false_α for some α . Then $\langle C, D \rangle$ is the smallest fixed point; that is, it is a fixed point included in every other fixed point. (We say $\langle A, B \rangle$ is included in $\langle C, D \rangle$ if $A \subseteq C$ and $B \subseteq D$.)

The existence of fixed points is startling. The liar paradox shows that, within the context of classical two-valued logic, it is not possible to assign an extension to ‘ Tr ’ in such a way that, for any sentence ϕ , ϕ is true or false according as $Tr(\ulcorner \phi \urcorner)$ is true or false. We can, however, solve the analogous problem for three-valued logic. At a fixed point, we assign an extension and an anti-extension to ‘ Tr ’ in such a way that, for any sentence ϕ , ϕ is true, false or unsettled according as $Tr(\ulcorner \phi \urcorner)$ is true, false or undecided.

There are fixed points other than the smallest. If A and B are non-overlapping sets with $\langle A, B \rangle \subseteq \Gamma(A, B)$, then there is a smallest fixed point which contains $\langle A, B \rangle$. The extension of ‘ Tr ’ in this smallest fixed point consists of the sentences which are derivable from the true atomic and negated atomic non-semantic sentences, together with the sentences in A and the negations of the sentences in B .

Fixed points other than the smallest are useful in describing the different sorts of pathological behaviour we find among self-referring sentences. Tarski’s sentence λ , which asserts its own untruth, is viciously paradoxical; whether we declare it true or false, we fall into contradiction (see Gödel’s theorems). By contrast, if we formalize the construction of the truth-teller sentence (‘The truth-teller sentence is true’), getting a sentence τ which is provably equivalent to $Tr(\ulcorner \tau \urcorner)$, we again get a sentence which is paradoxical, but in a more compliant fashion; whether we declare the sentence true or false, the sentence is amiably willing to go along, but no such classification has any basis in extra-semantic reality.

The difference between these two sorts of pathologies is reflected in the structure of the fixed points. There are no fixed points in which λ is assigned a truth-value. By contrast, there are some fixed points in which τ is true, others in which it is false, and still others in which it lacks a truth-value.

There are a great many fixed points; if the language has κ sentences, there will be 2^κ of them. The fixed points constitute a complete lower semi-lattice, that is, a partially ordered set in which every non-empty subset has a greatest lower bound. If $\mathcal{K}$ is a non-empty collection of fixed points, the union of all the fixed points which are included in every member of $\mathcal{K}$ will be a fixed point, so it will be the greatest lower bound of $\mathcal{K}$.

5 Classical-logic versions of Kripke's theory

Where Kripke's construction assigns a truth-value, the value it assigns invariably agrees with our intuitive understanding of the everyday application of 'true'. However, there are many sentences to which the construction fails to assign a truth-value but about which ordinary intuition fails to detect even a whiff of paradox. For example, 'A conjunction is true if both its conjuncts are true' is, according to intuition, harmlessly and obviously true, yet the construction fails to give it a truth-value. Indeed, the construction fails to provide even the simplest semantic generalizations, withholding its assent from such harmless platitudes as 'Every true sentence is a true sentence'; since $Tr(\ulcorner \lambda \urcorner)$ is truth-valueless, so is $(Tr(\ulcorner \lambda \urcorner) \rightarrow Tr(\ulcorner \lambda \urcorner))$, hence $(\forall x)(Tr(x) \rightarrow Tr(x))$ falls into the gap between truth and falsity.

This limitation is easily remedied by extending the rules of inference to encompass classical logic and taking as axiomatic the principle that, if a conditional and its antecedent are both true, the consequent is also true. Thus the variant construction treats a sentence as true just in case it is derivable from true atomic and negated atomic non-semantic sentences, together with instances of the schema

$$(Tr(\ulcorner \phi \rightarrow \psi \urcorner) \rightarrow (Tr(\ulcorner \phi \urcorner) \rightarrow Tr(\ulcorner \psi \urcorner))),$$

by ' Tr '-introduction, ' $\neg Tr$ '-introduction, universal quantification introduction, and the laws of classical logic.

According to this variant construction, all classically valid sentences are true, in particular, $(\forall x)(Tr(x) \rightarrow Tr(x))$. For each ϕ and ψ , the conditional

$$((Tr(\ulcorner \phi \urcorner) \wedge Tr(\ulcorner \psi \urcorner)) \rightarrow Tr(\ulcorner \phi \wedge \psi \urcorner))$$

is derivable from the fact that ' Tr ' distributes over conditionals, and the generalization that a conjunction is true if both its conjuncts are true follows by universal quantification introduction. Indeed, we can show that an inference schema is valid in classical logic if and only if all substitution instances of the schema are truth-preserving.

There is a price to pay. In the original construction, truth is unmistakably grounded in the non-semantic facts. In the alternative construction, this is not so evident, since we want an explanation of the truth of the laws of classical logic. In the original construction, the truth of a disjunction was based upon the truth of one of its disjuncts, but, in the alternative construction, this is no longer so.

Another construction with much the same effect employs Bas van Fraassen's method of 'supervaluations' (1966) in place of Kleene's three-valued logic; this is the method suggested by Kripke himself. The sentences declared true in $\Gamma(A, B)$ will be those sentences true in all the classical models we get from the original model by taking the extension of ' Tr ' to be a consistent set of sentences, closed under first-order consequence, which contains A and is disjoint from B ; the sentences declared false will be those whose negations are declared true. A sentence is regarded as true if it is declared true in the smallest fixed point.

The two constructions give similar results. Indeed, if the language is countable (which implies, since everything has a name, that the universe of discourse is countable), they give precisely the same results.

The classical-logic versions of Kripke's theory capture a great many of our common-sense intuitions about truth and paradox, but there remain fundamental intuitions that they leave aside. Notably, the intuition driving the theory is that the paradoxical sentences are neither true nor false. But the statement that the strengthened liar sentence is neither true nor false, while formalizable in the language, is never accepted as true. That the strengthened liar sentence is neither true nor false is something we can only see from the perspective of an essentially richer metalanguage. Within the object language, we can only recognize a sentence as untrue if at some

stage it is declared false; only from the external vantage point of the metalanguage can we regard a sentence as untrue because it is never declared true.

6 Revision theory

There are other proposals for obtaining an unambiguous truth predicate within a classical framework, notably the ‘revision theory’ of Hans Herzberger (1982), Anil Gupta (1982) and Nuel Belnap (1982), which sees the paradoxes as arising because the processes by which we ordinarily introduce new terms into the language, as seen in ordinary explicit definitions and in inductive definitions, partially break down when the (T)-sentences are taken to define ‘Tr’. The processes fail to stabilize upon a single choice for the extension of ‘Tr’. Instead, if V_0 is a candidate for the extension of ‘Tr’ we get another candidate, at least as good, by taking the extension of ‘Tr’ to be the set V_1 of sentences true in the classical model obtained by setting the extension of ‘Tr’ equal to V_0 . We get another good candidate by taking the extension of ‘Tr’ to be the set V_2 of sentences true in the model in which the extension of ‘Tr’ is V_1 . The process never settles down. Whereas Plato’s first statement above is in V_n for all $n > 5$, if the simple liar sentence is in V_n , it is outside V_{n+1} , but it is back inside V_{n+2} . The process continues into the transfinite, and, as it proceeds, we get better and better candidates for the extension of ‘Tr’. The process must eventually repeat itself, and, after it starts to do so, the candidates we get are all optimal candidates for the extension of ‘Tr’. Because the starting point V_0 was arbitrary, there are many different revision sequences. Count a sentence as ‘stably true’ if there is an ordinal level after which it always appears at every stage in every revision sequence.

The details of extending the process into the transfinite - what exactly to do at limit stages - is a matter of some disagreement among the three authors mentioned above, but all agree that a sentence is in V_λ , for λ a limit, if it is in all V_α , for α a sufficiently large ordinal less than λ , and a sentence is outside V_λ if it is outside V_α for sufficiently large $\alpha < \lambda$.

Revision theory still requires a richer metalanguage. We can only construct the revision sequences when the universe of discourse of the original language is a set, so we cannot carry out the construction for the language of set theory or for English, languages whose universe of discourse is not a set. (In those languages, we can talk about all sets, and there is no set that includes all sets.) The hope is that, in spite of this limitation, the fragments of English amenable to revision-theoretic treatment will be rich enough so that their study will cast a useful light upon our ordinary conception of truth in English.

7 English as an object language?

Most of the recent work on the paradoxes has been aimed at explicating the intuitions that govern ordinary speakers’ usage of the word ‘true’. For such purposes, a theory which only takes account of a fragment of the natural language may well suffice. For some theoretical purposes, however, one would like a global semantic theory, which takes full account of the linguistic practices of English speakers. Should it turn out, on the contrary, that a comprehensive theory of language is inherently impossible, so that part of our language lies irretrievably beyond the reach of linguistic inquiry, this would severely curtail the hopes of science, and it would also gravely undermine the prevalent mood of metaphysical naturalism, according to which human language is a part of the natural order, not intrinsically more mysterious or less comprehensible than the planetary orbits.

Hopes for a comprehensive theory are greatly discouraged by the liar paradox, but, if our optimism persists, we have only two alternatives: either to restrict classical logic or to restrict schema (T) (see §1 above). Restricting the logic does not seem to help. If we replace classical logic with Kleene’s three-valued logic, we still do not get the (T)-sentences. Indeed, Feferman’s investigations (1984) would seem to show that we do not get the (T)-sentences within any reasonable logic.

We might hope to preserve at least one direction of schema (T), but even here our hopes are thwarted by Montague’s theorem (1963) that any logical system that includes the left-to-right direction of (T) together with the ‘Tr’-introduction rule will be inconsistent with the laws of syntax. A dual result shows that the right-to-left direction is incompatible with ‘ $\neg Tr$ ’-introduction and also that it is incompatible with ‘ Tr ’-elimination, which is the converse of ‘Tr’-introduction.

Another possibility, recommended by Vann McGee (1991), is to replace schema (T) with a system of rules of

inference: ‘*Tr*’-introduction and -elimination and ‘ $\neg Tr$ ’-introduction and -elimination. The paradoxes are diagnosed as arising from the misapplication of these rules of inference within conditional proofs. The classical-logic versions of Kripke’s theory can be employed to show that this system of rules is consistent.

As this consistency proof shows, Kripke’s construction is a highly versatile tool, which can be usefully applied to a wide variety of philosophical purposes.

See also: [Inductive definitions and proofs](#); [Logical and mathematical terms, glossary of](#); [Paradoxes of set and property](#); [Tarski’s definition of truth](#); [Theory of types](#)

VANN MCGEE

References and further reading

- Barwise, J. and Etchemendy, J.** (1987) *The Liar*, New York: Oxford University Press. (A mathematically and philosophically sophisticated development of the thesis that the referent of ‘true’ depends on context, developed within situation semantics, as found in Barwise and Perry (1983).)
- Barwise, J. and Perry, J.** (1983) *Situations and Attitudes*, Cambridge, MA: MIT Press. (Situation semantics, utilized in Barwise and Etchemendy (1987).)
- Belnap, N.** (1982) ‘Gupta’s Rule of Revision Theory of Truth’, *Journal of Philosophical Logic* 11: 103-16. (Proposed modifying Gupta’s revision theory by changing the procedure for limit stages, a proposal accepted by Gupta.)
- Burge, T.** (1979) ‘Semantical Paradox’, *Journal of Philosophy* 76: 169-98; repr. in Martin (1984), 83-117. (Lucid and highly influential development of a context-dependent conception of truth.)
- Feferman, S.** (1984) ‘Toward Useful Type-Free Theories, I’, in Martin (1984), 237-87. (Careful investigation of a variety of unified approaches to semantic and set-theoretic paradoxes.)
- Fraassen, B.C. van** (1966) ‘Singular Terms, Truth Value Gaps, and Free Logic’, *Journal of Philosophy* 64: 464-95. (Invented supervaluations, used in §5.)
- Gödel, K.** (1931) ‘Über formal unentscheidbare Sätze der *Principia Mathematica* und verwandter Systeme I’, *Monatshefte für Mathematik und Physik* 38: 173-98; trans. ‘On Formally Undecidable Propositions of *Principia Mathematica* and Related Systems’, in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967, 592-617. (In proving the incompleteness of Peano arithmetic, Gödel developed the fundamental tools of modern logic.)
- Gupta, A.** (1982) ‘Truth and Paradox’, *Journal of Philosophical Logic* 11: 1-60; repr. in Martin (1984), 175-235. (Gupta and Herzberger independently discovered revision theory.)
- Gupta, A. and Belnap, N.** (1993) *The Revision Theory of Truth*, Cambridge, MA: MIT Press. (Extends the revision theory of truth, according to which ‘true’ is defined by the (T)-sentences, to an illuminating general theory of circular definitions.)
- Herzberger, H.** (1982) ‘Notes on Naïve Semantics’, *Journal of Philosophical Logic* 11: 61-102; repr. in Martin (1984), 133-74. (Herzberger and Gupta independently discovered revision theory.)
- Kleene, S.C.** (1952) *Introduction to Metamathematics*, New York: Elsevier, 1971. (An invaluable textbook. Three-valued logic is in §54.)
- Kripke, S.A.** (1975) ‘Outline of a Theory of Truth’, *Journal of Philosophy* 72: 690-716; repr. in Martin (1984), 53-81. (The centre of all contemporary discussion of the liar paradox.)
- Martin, R.** (ed.) (1984) *Recent Essays on Truth and the Liar Paradox*, New York: Oxford University Press. (A collection of the most important papers.)
- Martin, R. and Woodruff, P.** (1985) ‘On Representing “True-in-*L*” in *L*’, *Philosophia* 5: 213-17; repr. in Martin (1984), 47-51. (First proved the existence of fixed points, independently of Kripke.)
- McGee, V.** (1991) *Truth, Vagueness, and Paradox*, Indianapolis, IN: Hackett Publishing Company. (Explores prospects for a theory of truth developed without recourse to a richer metalanguage.)
- Montague, R.** (1963) ‘Syntactical Treatments of Modality, With Corollaries on Reflexion Principles and Finite Axiomatizability’, *Acta Philosophica Fennica* 16: 153-67. (A version of the liar paradox with ‘necessary’ in place of ‘true’.)
- Parsons, C.** (1974) ‘The Liar Paradox’, *Journal of Philosophical Logic* 3: 381-412; repr. in Martin (1984), 9-45. (Insightfully develops a unified approach to the semantic and set-theoretic paradoxes, based upon the context dependence of both the extension of ‘true’ and the domain of quantification.)
- Quine, W.V.** (1946) ‘Concatenation as a Basis for Arithmetic’, *Journal of Symbolic Logic* 11: 105-14. (Codes arithmetic into the theory of syntax.)

- Ramsey, F.P.** (1925) 'The Foundations of Mathematics', *Proceedings of the London Mathematical Society* 25: 338-84; repr. in *Philosophical Papers*, Cambridge: Cambridge University Press, 1990, 164-224. (Distinguishing semantic from set-theoretic paradoxes, Ramsey effects a great simplification in the theory of types by neglecting the former.)
- Tarski, A.** (1935) 'Der Wahrheitsbegriff in den formalisierten Sprachen', *Studia Logica* 1: 261-405; trans. J.H. Woodger (1956), 'The Concept of Truth in Formalized Languages', in *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*, ed. J. Corcoran, Indianapolis, IN: Hackett Publishing Company, 2nd edn, 1983, 152-278. (Virtually all modern work on the logic of truth takes this fundamental paper as its starting point.)
- Whitehead, A.N. and Russell, B.A.W.** (1910-13) *Principia Mathematica*, Cambridge: Cambridge University Press, 3 vols; 2nd edn, 1925-7; repr. 1994. (The development of all of mathematics on a single axiomatic basis. Discussion of the semantic paradoxes is in the introduction, 41-47 and 60-65.)

Semantics

Semantics is the systematic study of meaning. Current work in this field builds on the work of logicians and linguists as well as of philosophers. Philosophers are interested in foundational issues in semantics because these speak to the nature of meaning, as it embeds in our thinking and in our relations to each other and to the world. Of special interest are questions about how a semantic theory should respect the connections of meaning to truth and to understanding. In addition, numerous semantic problems concerning particular linguistic constructions bear philosophical interest, sometimes because the problems are important to resolving foundational semantical issues, sometimes because philosophical problems of independent interest are expressed using the constructions, and sometimes because clarity about the semantic function of the constructions enables clarity in the development of philosophical theories and analyses.

1 Historical sources and goals

Systematic theorizing about meaning in language always has been of philosophical interest. In addition to being a science with interesting foundational problems, semantics interests philosophers because of their special interests in the nature of meaning and in the meaning of particular sorts of language, notably the language that serves as our primary means of expressing philosophically interesting concepts and thoughts (see [Language, philosophy of](#)).

Contemporary semantics owes a great deal to work in formal logic by Gottlob [Frege](#), Bertrand [Russell](#) and Alfred [Tarski](#), among others (see [Logic, philosophy of](#)). This work has provided techniques for constructing *formal systems* - abstract systems of symbols and sets with rigorously defined logical and structural relationships (see [Formal languages and systems](#)). Ludwig [Wittgenstein](#) and logical positivists such as Rudolf [Carnap](#) found these techniques so compelling that they took certain formal, logical systems to reveal the kind of structure that any coherent language - in particular, any language suited to serious science - must possess at its roots (see [Logical positivism](#)). A less extreme but more enduring view is that many aspects of natural languages are aptly *modelled* by formal systems - whether formal systems devised independently by logicians or designed expressly to reveal the systematic structure of the languages in question.

Recently, semantics has profited from the development, under the aegis of Noam [Chomsky](#), of systematic accounts of natural language syntax (see [Syntax](#)). Chomsky regards the grammar of a person's language as a partly innate theory that is implicitly mastered in acquiring the language (see [Language, innateness of](#)). The linguist's task is to make this theory explicit, guided by the person's spontaneous judgments about which constructions are grammatically acceptable, by experimental evidence about their cognitive faculties, and by general evidence about innate constraints on human grammars. Armed with this conception of the linguist's task, Chomsky and others have developed powerfully sophisticated theories of natural language grammar. Contemporary semanticists find this work useful for a number of reasons. Many semanticists now view their task in a Chomskian light: to uncover partly innate, cognitively real theories. Also, most contemporary semantical theories begin with something very like a grammar: a systematic account of the way sentences of the language in question are constructed from their meaningful parts - and it is quite reasonable to expect that the semantically relevant kind of sentence-structure bears a close relation to the correct grammar for the language.

2 Contemporary semantical traditions

The meaning of a statement is a matter of how things are, according to the statement; and the meaning of any expression is what users attach to it in their competent understanding of it. Semantics, then, ought to respect the outward-looking aspect of meaning: meaning fixes the representational and descriptive powers of language. And semantics must also respect the cognitive and epistemological aspect of meaning: meaning is cognized - it is grasped, known, understood.

A language (whether it be a shared social entity such as English, a narrower dialect, or even an individual's own 'idiolect') contains reusable words and expressions such as words and sentences. These are called expression-types because different tokenings of the expressions (different utterances or inscriptions of them) are importantly similar - they are of the same type (see [Type/token distinction](#)). Expression-types in a language have 'timeless' meaning that is drawn on in the various tokenings of the expressions. However, meanings can be expressed in the tokenings

that complete, or go beyond, the timeless meanings of the expressions used. For instance, a particular utterance of the sentence ‘That was worse’ would express a meaning to do with a particular subject matter (‘that’) and would involve (because of ‘was worse’) a gesture to another subject matter and to a dimension of evaluation. An attractive principle is that the timeless meaning of an expression simply is a matter of the expression’s distinctive potential for expressing meanings in particular tokenings. The discipline of pragmatics (whose border with semantics is notoriously porous) seeks to understand the relations between the timeless meaning of expression-types and the meanings achieved by tokenings, as well as the variety of ways tokenings (or speech acts) can achieve communicative and other purposes (see [Pragmatics](#); [Presupposition](#); [Propositions, sentences and statements](#); [Speech acts](#)).

A fact about natural languages that is no less impressive for being obvious is that every natural language issues in infinitely many sentences with different meanings. In English, ‘Ann ate a pear’ differs in meaning from ‘Ann ate a pear, and then Ann ate a pear’, which differs too from ‘Ann ate pear, and then Ann ate a pear, and then Ann ate a pear’, and so on. Any acceptable semantic account of a natural language will reveal its meanings as depending systematically on its repeatable, recombinant features (features of its expression-types and of their tokenings). One way to accomplish this is to show how meanings of complex expressions are generated from the meanings of simpler expressions, according to how the simpler expressions are combined. Such a strategy might employ a notion of *semantic values* of the meanings that are assigned to the simpler expressions (or tokenings) in the generation of the meanings of the more complex. That structure of a linguistic expression which, on these accounts, provides the key to the compositional derivation of its meaning sometimes is called the expression’s *logical form*, because of assumptions about the nature of meaning to which we now turn (see [Compositionality](#); [Logical form](#)).

The representational side of meaning has led many semanticists to take as their primary goal systematic accounts of the *truth-conditions* of sentences, as well as accounts of such related phenomena as the conditions of true application of predicates (such as ‘is a horse’) and the conditions of the denotation or reference of various noun phrases (such as proper names, definite descriptions like ‘the last Tsar’, and demonstratives like ‘this’). It is immediately plausible that conditions of truth, applicability and reference are proper subjects for a semantic theory. For instance, the meaning of ‘is a felucca’, very plausibly, is a matter of what it is for a thing to be a felucca, which, again plausibly, is given by revealing what it takes for a thing to be truly described as a ‘felucca’. Similarly, the meaning of ‘In Syracuse there are many feluccas’ is elucidated by characterizing how things would have to be to be truly so described (see [Meaning and truth](#)).

Because of the widespread view that a complete semantic account of a language must include systematic explanations of what it takes for its expressions to be true, to refer, or to apply truly, the central foundational issues in semantics include the nature of truth and reference, as well as the proper explanation and use within a semantic theory of the notions of truth- and applicability-conditions. Since truth- and application-conditions stand to each other in logical relations such as consistency and entailment, accounts of these conditions can serve what is regarded as another important goal in semantics: to explain the logical relations among linguistic expressions (and among their tokenings) (see [Truth, coherence theory of](#); [Truth, correspondence theory of](#); [Truth, deflationary theories of](#); [Truth, pragmatic theory of](#); [Semantic paradoxes and theories of truth](#); [Reference](#); [Sense and reference](#)).

A style of contemporary semantics that can be traced to Frege and Russell employs as semantic values abstract ‘universals’ such as *propositions* and *properties* - entities that embody truth- or application-conditions, in that (unlike linguistic expressions) they are not intrinsically tied to any particular language and have their truth- or application-conditions timelessly and essentially. Thus, some semantical frameworks start with conceptions of the nature of propositions and properties; within these frameworks the principal task in giving a semantical description of a given language is to discover the principles governing which expression-tokenings have as their semantic values which propositions and properties (see [Propositions, sentences and statements](#); [Universals](#); [Intensional entities](#)).

Some of these frameworks employ devices from the logical theory of models, in which the meanings of expressions are represented by set-theoretical structures (see [Model theory](#)). A particularly influential tradition in semantics deriving from the work of Carnap, Saul Kripke and Richard Montague, uses or imitates the notion of a ‘possible world’: the meaning of a statement is given by the sets of possibilities it excludes and allows, and the

meanings of smaller expressions are their contributions to this semantic function of the sentences in which they figure (see [Semantics, possible worlds](#); Carnap, R.; Montague, R.M.; Kripke, S.A.; [Semantics, situation](#)). Among the philosophical insights claimed for this work is the idea that in addition to revealing a statement's simple truth-condition (the condition that the utterance must meet if it is to be true), a semantical theory ought further to reveal its *modal content*, which determines what sorts of possible situation the utterance correctly describes. Focus on this distinction has promoted intense debate about metaphysical necessity and possibility, in particular, on whether certain devices of natural language (such as proper names) presuppose substantive views about necessity and essence, and on whether these views are seriously tenable (see [Proper names](#); [Essentialism](#); Kripke, S.A.).

A somewhat different tradition in semantics has been pioneered by Donald [Davidson](#), employing techniques due to Tarski (see [Tarski's definition of truth](#)). The goal in offering a Davidsonian semantics for a given language is to find an appropriate 'truth theory', thought of as a deductive system whose axioms and rules issue in 'T-sentences' such as 'The sentence "Schnee ist weiss" is true if and only if snow is white'. An appropriate truth-theory will not only be correct and complete, but also will in some sense explain the language-user's competence by revealing what it is that the speaker knows in knowing the language (see [Meaning and truth](#); [Meaning and understanding §2](#)).

Recently, too, philosophers, linguists and cognitive scientists have developed systematic semantic accounts based on algorithmic or procedural conceptions of meaning. These traditions borrow from the logical theory of proof (see [Semantics, game-theoretic](#); [Proof theory](#)).

All semantical traditions, even those that focus heavily on truth-conditions, must aim as well at least for compatibility with satisfactory explanations of the cognitive/epistemological aspects of meaning, including understanding (see [Meaning and understanding](#); [Meaning and verification](#); [Sense and reference](#)). One strategy is to keep issues about understanding at a remove from detailed semantic theorizing, by viewing understanding as involved not so much in the nature of meaning as in the uses to which meaning is put in thought and communication. David Lewis, for instance, proposes that a semantic account of a language is simply a correct description of the abstract function that assigns to statements in the language sets of possible worlds as meanings; understanding enters the picture only when we consider what constitutes the language's actually being used. For Lewis, the competent use of a particular language amounts to participation in a certain sort of conventional practice (see [Language, conventionality of](#); Lewis, D.K.). Other strategies take semantic theorizing to be more directly focused on what the competent understanding of a language consists in. Some accounts in the Davidsonian tradition, for instance, submit truth-theories to constraints of cognitive plausibility: a truth-theory must not employ conceptual resources beyond those plausibly required of a competent language-user. Other accounts allow alien concepts in a semantic description of a language and employ a notion of *tacit knowledge* to explain how the semantic description represents what the language-user knows (see [Knowledge, tacit](#)). Also heavily cognitive in orientation are accounts within the Chomskian paradigm; these take the primary goal of semantics to be the discovery of the cognitive capacities and structures underlying linguistic competence (see [Chomsky, N.](#)). Some semantic accounts take as their task the explanation of how linguistic meaning is fixed by the meaning of mental items such as concepts and thoughts; for the ultimate source of meaning, these accounts defer to theories of cognitive semantics (see [Mind, philosophy of](#); Grice, H.P.; [Communication and intention](#); [Concepts](#); [Language of thought](#); [Semantics, conceptual role](#); [Semantics, informational](#); [Semantics, teleological](#)).

3 Issues and problems in semantics

Some debates about the semantic features of particular sorts of expression arise only within particular semantic traditions. Questions of *reference*, in contrast, are of very broad interest, since reference is widely taken to be a semantically central notion. Philosophers have hotly pursued questions about the nature of reference, about the proper role of reference in semantic theories and about how the reference of particular kinds of linguistic item is determined (see [Reference](#); [Sense and reference](#)). Receiving special attention have been proper names, natural kind terms (like 'tiger' and 'gold'), context-dependent terms (such as the demonstrative 'that' and the indexical 'today') and theoretical terms of science (such as 'mass' and 'oxygen') (see [Proper names](#); [Natural kinds](#); [Demonstratives and indexicals](#); [Scientific realism and antirealism](#)). Among the questions under heavy discussion are the extent to which the determinants of reference are settled by intrinsic features of a speaker's cognition, rather than being settled by features of the speaker's society and environment (see [Language, social nature of](#); [Content: wide and](#)

narrow; Methodological individualism).

Issues about the semantical or logical structure of various locutions, in contrast, are often discussed in the context of particular traditions of semantic theory (see **Logical form**). Still, a number of such issues have interest that is not confined to any one semantic tradition. Among these are the proper treatment of definite descriptions such as ‘the short spy’. Definite descriptions resist easy assimilation either to the paradigm of proper names or to that of quantifying phrases such as ‘all circles’ and ‘three dogs’ (see **Descriptions; Quantifiers**). Also of persisting interest is the behaviour of ‘anaphoric’ pronouns - ones which in one way or another ‘refer back’ (or even ahead) - such as occur in ‘The dog is happy; he is wagging his tail’ and ‘Probably some dogs can climb trees; it is likely that they are small’. Difficulties in providing systematic explanations of the roles of anaphoric pronouns has led to a number of theories according to which semantic accounts of sentences must view their semantically relevant structure as inextricable from the larger discourse in which the sentences figure (see **Discourse semantics; Anaphora**). Even the logical form of apparently simple subject-predicate sentences has been debated. For instance, Davidson proposes that the logical form of ‘Bill is running’ involves not the simple predication of an attribute to Bill, but rather an existence claim about an event (perhaps: an event exists, which is a running, which is in progress, and of which Bill is the agent). This account is meant to cohere with a straightforward account of adverbs, in which, for instance, it is clear why ‘Bill is running’ is entailed by ‘Bill is running slowly’ (see **Predication; Events; Adverbs**).

So-called propositional-attitude ascribing statements (such as ‘Sue believes that oranges are expensive’) also have prompted extensive discussion. An ancient puzzle still has received no widely accepted solution: how is it that Hesperus and Phosphorus are one and the same entity, and yet ‘Hammurabi believed that Hesperus is bright’ might be true while ‘Hammurabi believed that Phosphorus is bright’ is false (see **Propositional attitude statements; Indirect discourse; Intensionality**)?

Another source of fascinating, but inconclusive, debate among semanticists is the treatment of conditionals (if/then sentences), including indicative conditionals like ‘If Oswald didn’t shoot Kennedy, then someone else did’ and subjunctive, counterfactual conditionals like ‘If Oswald hadn’t shot Kennedy, then someone else would have’. Concerning indicative conditionals, one current view is that their truth-conditions are aptly modelled on those of the material conditional of formal logic, while another is that indicative conditionals have no truth-conditions at all but only conditions of proper assertion (see **Indicative conditionals**). Counterfactual conditionals have been linked to metaphysical issues about necessity, causation and natural law, and debates about their interpretation are enmeshed with metaphysical debates (see **Counterfactual conditionals**).

Also among the issues that bear on semantic theories are questions about necessity and possibility, the nature and ontological acceptability of intensional entities, tense, metaphor, vagueness and language apparently about nonexistent and fictional entities (see **Modal operators; Modal logic, philosophical issues in; Intensional logics; Intensional entities; Tense and temporal logic; Metaphor; Vagueness; Existence; Fiction, semantics of; Fictional entities; Necessary truth and convention**).

See also: **Ambiguity; Analyticity; Emotive meaning; Imperative logic; Intuitionistic logic and antirealism; Logical constants; Mass terms; Meaning, Indian theories of; Meaning in Islamic philosophy; Ontological commitment; Questions; Semiotics; Structuralism in linguistics**

MARK CRIMMINS

References and further reading

- Larson, R. and Segal, G.** (1995) *Knowledge of Meaning*, Cambridge, MA: MIT Press. (A lucid and philosophically sophisticated introduction to semantics developing a broadly Davidsonian approach.)
- Chierchia, G. and McConnell-Ginet, S.** (1995) *Meaning and Grammar*, Cambridge, MA: MIT Press. (An excellent introduction to semantics in the tradition of Montague grammar.)

Sense and reference

The *'reference'* of an expression is the entity the expression designates or applies to. The *'sense'* of an expression is the way in which the expression presents that reference. For example, the ancients used *'the morning star'* and *'the evening star'* to designate what turned out to be the same heavenly body, the planet Venus. These two expressions have the same reference, but they clearly differ in that each presents that reference in a different way. So, although coreferential, each expression is associated with a different *'sense'*. The distinction between sense and reference helps explain the cognitive puzzle posed by identity statements. *'The morning star is the evening star'* and *'The morning star is the morning star'* are both true, yet the sentences differ in cognitive significance, since the former may be informative, whereas the latter definitely is not. That difference in cognitive significance cannot be explained just by appeal to the references of the terms, for those are the same. It can, however, be naturally accounted for by appeal to a difference in sense. The terms *'the morning star'* and *'the evening star'* used in the first sentence, having different senses, present the referent in different ways, whereas no such difference occurs in the second sentence.

The distinction between sense and reference applies to all well-formed expressions of a language. It is part of a general theory of meaning that postulates an intermediate level of sense between linguistic terms and the entities the terms stand for. Senses give significance to expressions, which in and of themselves are just noises or marks on a surface, and connect them to the world. It is because linguistic terms have a sense that they can be used to express judgments, to transmit information and to talk about reality.

1 Sense, reference and cognitive significance

The distinction between sense and reference was originally drawn by Gottlob Frege in his 1892 article *'Über Sinn und Bedeutung'* ('On Sense and Reference'; see Frege, G. §3). According to Frege, associated with each meaningful expression of a language there is a *'sense'* (*Sinn*) that makes the expression significant and determines its *'reference'* (*Bedeutung*), that is, the entity the expression applies to or designates. (*Bedeutung* is also translated as *'designation'*, *'designatum'*, *'denotation'* and *'nominatum'*. It is also sometimes translated as *'meaning'*. This can create confusion, since some of Frege's contemporaries, including Russell, used *'meaning'* as the translation of *'Sinn'*, most likely on the assumption that the notion corresponding to *'Sinn'* is that of cognitive meaning or significance.) Before 1892, Frege had espoused the view that the significance or content of an expression is given exclusively by that which the expression designates or stands for (what he came to regard as the *'reference'*). Observe that on such a view two coreferential terms, such as *'Cicero'* and *'Tully'*, are supposed to have the same content or significance. But if that is so, it seems impossible to explain the obvious difference in cognitive significance between, say, *'Cicero is Cicero'*, a trivially true and uninformative statement, and *'Cicero is Tully'*, a potentially informative sentence. The problem of how to explain differences in cognitive significance is not restricted exclusively to statements of identity. *'Cicero was an orator'* and *'Tully was an orator'* also differ in cognitive significance, since it is possible for a competent speaker to accept one of the statements while rejecting the other. Without postulating a difference in the content or significance of the terms *'Cicero'* and *'Tully'*, it would seem impossible to explain the differences between those two pairs of sentences. The distinction between sense and reference provides for a natural explanation: although the two names are coreferential, they are associated with different senses and those senses present the referent in different ways (see Proper names §1).

2 The sense and reference of different types of expressions

The distinction between sense and reference is intended to apply to all meaningful expressions of a language, whether simple or complex. In the case of definite descriptions it is an extremely natural distinction since it is obvious that different descriptions of one and the same object present that object in different ways. In the case of proper names, the reference is the bearer of the name and the sense is traditionally conceived as some sort of condition satisfied uniquely by the referent. There is no unanimity, however, regarding how to understand such a uniquely identifying condition. Frege suggests (1892) that the sense of a proper name is some descriptive information that applies uniquely to the bearer. Thus, some philosophers interpret Frege's suggestion as entailing the claim that each proper name is synonymous with some definite description. Others, including John Searle (1958), view the sense of a given proper name as a cluster of descriptive information that speakers associate with

the name, and the referent as whatever satisfies most of the attributes in the cluster. Other semanticists do not uphold the descriptive nature of the senses of proper names. For example, David Kaplan (1969) has argued that on Frege's view the sense of a name may be seen as a representation of the referent and so the relation between the sense of a name and its reference is much like the relation between picture and pictured object. Finally, Michael Dummett (1981) is the foremost representative of the view that regards the sense of a proper name, as the sense of any expression, as a mechanism or procedure that, under ideal circumstances, would enable a competent speaker to single out the referent.

The reference of a general term is traditionally taken to be the extension of the term in question. Thus, for example, in the case of a predicate such as 'red' the extension is the set of red things, that is, the set of things to which the predicate correctly applies. And the extension of a relational verb such as 'is taller than' is a set of ordered pairs (since 'being taller than' is a relation between two things) such that the first member of each pair is taller than the second member of the pair. The sense of a general term is a mode of presentation of the set and is responsible for the cognitive content or cognitive significance of the term.

Although it is traditional to represent the reference of a general term as a set, such a choice is strictly incorrect, from the standpoint of pure Fregean orthodoxy. Frege distinguished sharply between concepts and objects. Only the latter are, on his view, 'saturated' entities; entities that subsist on their own. People and artefacts, also numbers and sets, are saturated entities. On the other hand, concepts are 'unsaturated'; they do not have an independent existence, but can only subsist as completed by a saturated entity. On Frege's view, the reference of a general term is a concept, and thus an unsaturated entity. Thus, the reference of a general term is not a set, for the latter is a saturated entity. Still, in the posthumously published comments on '*Über Sinn und Bedeutung*' (1892-5), Frege himself stresses that any two general terms refer to the same concept just in case their extension is the same. So, it seems that, even for Frege, equality of extension is a criterion of identity for concepts, so taking the reference of a general term to be a set is not a radical departure from the standard Fregean theory.

Finally, according to the theory, whole sentences express complete thoughts. The 'thought' or sense associated with a sentence is the claim or assertion, traditionally called the 'proposition', expressed by the sentence. In '*Über Sinn und Bedeutung*' Frege argues that the sentence's reference is its truth-value. The relation between a sentence and its truth-value should therefore be understood in terms of the relation between a name and its bearer or a definite description and the individual it designates. A true sentence is like a name for 'True', so that all the true sentences (like all the false ones) have one and the same reference; however, since different, non-synonymous sentences express different thoughts, they present that reference in different ways. Frege's argument relies on the assumption (see §4 below) that the reference of a complex expression does not change if components with the same reference are intersubstituted. Thus, for example, if we substitute the coreferential 'the orator who denounced Catiline' for 'Cicero' in the complex expression 'the father of Cicero', the reference of the expression (whoever fathered the famous orator) does not change. In this vein Frege asks what it is that remains constant when we intersubstitute coreferential components in a sentence, for that should be the reference, if anything is. He observes that the thought cannot be the reference, for typically the thought expressed by the sentence is altered: 'Cicero was an orator' and 'The orator who denounced Catiline was an orator' clearly express different thoughts. The truth-value of the sentence is what remains constant in spite of the substitution, so that should be the reference.

3 The multiple roles of sense

The sense of an expression, as we have seen, embodies its cognitive value and is solely responsible for its significance. A term that lacks a sense is therefore not a meaningful constituent of a language; however, an expression may have a sense (that is, be meaningful) and still lack reference. For example, 'the least integer' is a meaningful expression even if there is no object to which it refers. The sense of an expression determines the referent (if the expression has a referent) but, as Frege points out, there is no road back from reference to sense. Aristotle, for example, is the referent of many different terms; there is no mechanism, no procedure that would single out a sense only on the basis of what the reference is.

The sense of a sentence - the thought or proposition it expresses - determines a truth-value as reference. But truth-value typically depends on specific features of the world. Thus, given the way the world is, the sense of the sentence 'Aristotle was a philosopher' determines (we suppose) 'True' as reference. Had Aristotle never thought about philosophical questions, the sense of that sentence would determine 'False' as reference. It is the role of the

sense of a sentence to determine truth-value on whatever conditions obtain, thus the sense accounts for the truth-conditions of a sentence. Because the thought expressed by a sentence is a function of the senses of the component expressions, the sense of an expression is that expression's contribution to the truth-conditions of the sentences in which it figures.

Senses are also what speakers grasp when they understand words. Hence, language learning and linguistic competence are a matter of grasping senses. Moreover, speakers communicate linguistically with one another by grasping the same thoughts. So senses, although mentally graspable entities, should not be confused with private images or ideas. Senses are objective, so that different speakers can grasp the same senses and transmit a given thought to other speakers.

4 Compositionality and substitutivity

The Fregean theory of sense and reference espouses two basic and closely connected semantic principles: the principle of compositionality and the law of substitutivity.

The principle of compositionality states that the reference of a complex expression is a function of the references of its components, plus the way in which those components are combined (see **Compositionality**). Similarly, the sense of a complex expression is a function of the senses of the component expressions. Thus the reference of the sentence 'The Eiffel Tower is taller than the Empire State Building', a truth-value, is determined exclusively by the references of the meaningful components and their combination. The mode of combination is important, since a different reference would be determined if the components were combined differently, as with the sentence 'The Empire State building is taller than the Eiffel Tower'.

Since, by compositionality, the references of the parts are solely responsible for the reference of the whole, it obviously follows that coreferential terms are intersubstitutable without altering the reference of the complex expressions in which they occur. The latter is known as the law of substitutivity, which states that an expression which figures as part of another expression can be replaced with a coreferential one, and the reference of the complex will remain the same. However, the *sense* of a complex is typically altered when coreferential components are intersubstituted, if they have different senses. Thus, for example, 'Aristotle was Greek' and 'The tutor of Alexander the Great was Greek' express different thoughts but, since the latter sentence results from the exchange of coreferential terms, the reference of the two sentences is the same. (The sense of a complex remains the same if components with the same sense are intersubstituted.)

5 Non-customary contexts

The law of substitution seems to have obvious counterexamples. For example, although 'Cicero' and 'Tully' are coreferential names, the reference of sentences such as 'Mary said that Cicero was a historian' or 'John said, "Cicero was a historian"' may be altered if 'Tully' is substituted for 'Cicero'. But, according to the standard theory of sense and reference, some contexts cause shifts in the sense and reference of the expressions that occur in them. Thus the alleged failure of substitutivity is only apparent, for in the contexts in question the terms being substituted do not have their customary reference. The Fregean theory of sense and reference is essentially contextual: what an expression refers to and what its sense is depend on the context in which the expression occurs.

Consider the sentence 'The word "Cicero" contains six characters'. The reference of a word within quotation marks is not the customary referent, but rather the word itself (see **Use/mention distinction and quotation**). In our example, the occurrence of 'Cicero' does not refer to Cicero but to the word 'Cicero'. Clearly, the substitution of 'Tully' for 'Cicero' in that context would alter the truth-value of the sentence in question. But since 'Tully' and 'Cicero' are different words, the terms of the attempted substitution are not coreferential in that context, and therefore the observed phenomenon does not violate the principle of substitutivity.

Something similar occurs in the case of indirect quotation, such as 'Mary said that Cicero was a historian' (see **Indirect discourse**), and belief ascriptions, such as 'John believes that Cicero was a historian' (see **Propositional attitude statements §§1-2**). Here too the intersubstitution of what seem to be coreferential terms may alter the reference of the sentence. And here again, the breakdown of substitutivity is only apparent, for, according to the theory of sense and reference, expressions in contexts such as these, which Frege called 'oblique', have a non-customary referent. The non-customary referent in these cases - the oblique referent - is the customary sense

of the expression. In oblique contexts the reference of the embedded clause is the thought which that clause customarily expresses, and the reference of each of the clause's component expressions is its customary sense. So, the reference of 'Cicero' in the two sentences considered above is not Cicero but rather the sense 'Cicero' has in customary contexts such as 'Cicero was an orator', 'Cicero was a historian' and so on. In oblique contexts expressions have in turn an oblique sense, whose role is precisely to determine the oblique reference. It is an open question whether the semantic treatment of more complex embeddings requires that expressions have a higher-order sense that determines the oblique sense as reference in the context in question.

See also: [Intensional entities §1](#)

GENOVEVA MARTÍ

References and further reading

- Burge, T.** (1990) 'Frege on Sense and Linguistic Meaning', in D. Bell and N. Cooper (eds) *The Analytic Tradition: Meaning, Thought and Knowledge*, Philosophical Quarterly Monographs series, vol. 1, Oxford and Cambridge, MA: Blackwell, 30-60. (Proposes a nontraditional interpretation of the notion of sense.)
- Church, A.** (1951) 'A Formulation of the Logic of Sense and Denotation', in P. Henle, H. Kallen and S. Langer (eds) *Structure, Method and Meaning*, New York: Liberal Arts, 3-24. (A formal development of Frege's ideas. Church argues for the need for an infinite hierarchy of senses.)
- Donnellan, K.S.** (1972) 'Proper Names and Identifying Descriptions', in D. Davidson and G. Harman (eds) *Semantics of Natural Language*, Dordrecht: Reidel, 356-79. (A criticism of the Fregean theory of sense and reference as applied to proper names.)
- Dummett, M.** (1981) *Frege: Philosophy of Language*, Cambridge, MA: Harvard University Press. (A comprehensive discussion of various aspects of the theory of sense and reference.)
- Evans, G.** (1982) *The Varieties of Reference*, Oxford: Clarendon Press. (A contemporary Fregean approach to sense and reference. Chapter 1 has a thorough exposition and discussion of the doctrine of sense and reference for proper names.)
- Evans, G.** (1985) 'Understanding Demonstratives', in *Collected Papers*, Oxford: Oxford University Press. (Applies the theory of sense and reference to indexicals.)
- Frege, G.** (1879) *Begriffsschrift, eine der arithmetischen nachgebildeten Formelsprachen des reinen Denkens*, Halle: Nebert; trans. 'Conceptual Notation', in *Conceptual Notation and Related Articles*, ed. T. Bynum, Oxford: Clarendon Press, 1972, 101-203. (A sample of Frege's thought before the distinction between sense and reference. §8 gives Frege's first attempted solution to the problem posed by true and informative identity sentences.)
- Frege, G.** (1892) 'Über Sinn und Bedeutung', *Zeitschrift für Philosophie und philosophische Kritik* 100: 25-50; trans. and ed. P.T. Geach and M. Black, 'On Sense and Meaning', in *Translations from the Philosophical Writings of Gottlob Frege*, Oxford: Blackwell, 3rd edn, 1980, 56-78; trans. H. Feigl, 'On Sense and Nominatum', in H. Feigl and W. Sellars (eds) *Readings in Philosophical Analysis*, New York: Ridgeview, 1982, 85-102. (This is the essay in which Frege introduces the distinction between sense and reference.)
- Frege, G.** (1892-5) 'Ausführungen über Sinn und Bedeutung', in *Nachgelassene Schriften*, ed. H. Hermes, F. Kambartel and F. Kaulbach, Hamburg: Meiner, 1969-76; trans. P. Long and R. White, 'Comments on Sense and Meaning', in *Posthumous Writings*, ed. H. Hermes, F. Kambartel and F. Kaulbach, Chicago, IL: Chicago University Press, 1979, 118-25. (A posthumously published discussion of some of the major themes of 'Über Sinn und Bedeutung'. Frege focuses specially on the application of the notions of sense and reference to general terms.)
- Frege, G.** (1918) 'Der Gedanke: eine logische Untersuchung', *Beiträge zur Philosophie des deutschen Idealismus* 1: 58-77; trans. A.M. and M. Quinton, 'The Thought: A Logical Inquiry', *Mind* 65: 289-311; trans. P.T. Geach and R. Stoothoff, 'Thoughts', in *Collected Papers*, ed. B. McGuinness, Oxford: Blackwell, 1984. (Frege's defence of the objectivity of the realm of sense. It also includes a few remarks on the sense of indexicals.)
- Furth, M.** (1968) 'Two Types of Denotation', *Studies in Logical Theory*, American Philosophical Quarterly Monograph series, vol. 2, Oxford: Blackwell, 9-45. (Discusses unsaturated entities and the reference of general terms.)
- Kaplan, D.** (1969) 'Quantifying In', in D. Davidson and J. Hintikka (eds) *Words and Objections*, Dordrecht: Reidel, 178-214. (A discussion of Frege's treatment of oblique contexts. Argues for the representational conception of the sense of proper names.)
- Parsons, T.** (1981) 'Frege's Hierarchies of Indirect Senses and the Paradox of Analysis', in P.A. French, T.E. Uehling, Jr and H.K. Wettstein (eds) *Midwest Studies in Philosophy*, vol. 6, 37-57. (Contends, against Church,

that just one level of oblique sense and reference is sufficient to provide an adequate semantic treatment of all contexts.)

Perry, J. (1977) 'Frege on Demonstratives', *Philosophical Review* 86: 474-97. (Criticizes the theory of sense and reference as applied to indexicals.)

Salmon, N. (1986) *Frege's Puzzle*, Cambridge, MA: MIT Press. (A critical discussion of the puzzle of cognitive significance.)

Searle, J.R. (1958) 'Proper Names', *Mind* 67: 166-73. (On the cluster conception of the sense of proper names.)

Sluga, H. (ed.) (1993) *The Philosophy of Frege*, vol. 4, *Sense and Reference in Frege's Philosophy*, New York: Garland. (Large collection of articles about the topic.)

Set theory

In the late nineteenth century, Georg Cantor created mathematical theories, first of sets or aggregates of real numbers (or linear points), and later of sets or aggregates of arbitrary elements. The relationship of element a to set A is written $a \in A$; it is to be distinguished from the relationship of subset B to set A , which holds if every element of B is also an element of A , and which is written $B \subseteq A$. Cantor is most famous for his theory of transfinite cardinals, or numbers of elements in infinite sets. A subset of an infinite set may have the same number of elements as the set itself, and Cantor proved that the sets of natural and rational numbers have the same number of elements, which he called $\aleph_0$; also that the sets of real and complex numbers have the same number of elements, which he called c . Cantor proved $\aleph_0$ to be less than c . He conjectured that no set has a number of elements strictly between these two.

In the early twentieth century, in response to criticism of set theory, Ernst Zermelo undertook its axiomatization; and, with amendments by Abraham Fraenkel, his have been the accepted axioms ever since. These axioms help distinguish the notion of a set, which is too basic to admit of informative definition, from other notions of a one made up of many that have been considered in logic and philosophy. Properties having exactly the same particulars as instances need not be identical, whereas sets having exactly the same elements are identical by the axiom of extensionality. Hence for any condition Φ there is at most one set $\{x|\Phi(x)\}$ whose elements are all and only those x such that $\Phi(x)$ holds, and $\{x|\Phi(x)\} = \{x|\Psi(x)\}$ if and only if conditions Φ and Ψ hold of exactly the same x . It cannot consistently be assumed that $\{x|\Phi(x)\}$ exists for every condition Φ . Inversely, the existence of a set is not assumed to depend on the possibility of defining it by some condition Φ as $\{x|\Phi(x)\}$.

One set x_0 may be an element of another set x_1 which is an element of x_2 and so on, $x_0 \in x_1 \in x_2 \in \dots$, but the reverse situation, $\dots \in y_2 \in y_1 \in y_0$, may not occur, by the axiom of foundation. It follows that no set is an element of itself and that there can be no universal set $y = \{x|x = x\}$. Whereas a part of a part of a whole is a part of that whole, an element of an element of a set need not be an element of that set.

Modern mathematics has been greatly influenced by set theory, and philosophies rejecting the latter must therefore reject much of the former. Many set-theoretic notations and terminologies are encountered even outside mathematics, as in parts of philosophy:

pair	$\{a, b\}$	$\{x x = a \text{ or } x = b\}$
singleton	$\{a\}$	$\{x x = a\}$
empty set	$\emptyset$	$\{x x \neq x\}$
union	$\cup X$	$\{a a \in A \text{ for some } A \in X\}$
binary union	$A \cup B$	$\{a a \in A \text{ or } a \in B\}$
intersection	$\cap X$	$\{a a \in A \text{ for all } A \in X\}$
binary intersection	$A \cap B$	$\{a a \in A \text{ and } a \in B\}$
difference	$A - B$	$\{a a \in A \text{ and not } a \in B\}$
complement	$A - B$	
power set	$\wp(A)$	$\{B B \subseteq A\}$

(In contexts where only subsets of A are being considered, $A - B$ may be written $-B$ and called the complement of B .)

While the accepted axioms suffice as a basis for the development not only of set theory itself, but of modern mathematics generally, they leave some questions about transfinite cardinals unanswered. The status of such questions remains a topic of logical research and philosophical controversy.

1 The axioms of set theory

The axiom of extensionality states that sets having exactly the same elements are identical. The axiom of foundation states that there is no infinite descending sequence of sets $\dots \in x_2 \in x_1 \in x_0$. The basic

Zermelo-Fraenkel axioms (ZF; see [Zermelo 1908](#)) consist of the axioms of extensionality and foundation and existence axioms stated below. The accepted axioms ZFC consist of ZF plus the axiom of choice (AC): given a partition or set x of sets that are non-empty (each has at least one element) and disjoint (no two have a common element), there exists a choice set y having exactly one element in common with each element of x .

The intuition behind the axioms is that sets lie in a hierarchy of levels (see [Set theory, different systems of §5](#)), and when $x \in y$, then x lies below and y , above. At the lowest level there is just $\emptyset$ (though a variant of the standard notion allows at this level ‘individuals’ or items that have no elements and are not sets but may be elements of sets of any level). There is no highest level (though a variant allows a top level of ‘classes’ or items that may have sets of any level as elements but are not themselves sets nor elements of anything). If set x lies at one level then the ‘successor’ of x , $x^\# = x \cup \{x\}$ lies at the next higher level. The levels continue into the infinite, so that above $\emptyset$, $\emptyset^\# = \{\emptyset\}$, $\emptyset^{\#\#} = \{\emptyset, \{\emptyset\}\}$, ... lies the set $\{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}, \dots\}$, whose existence is asserted by the axiom of *infinity*. The hierarchy is supposed to be as ‘tall’ and as ‘wide’ as possible, an intuition (only) partly expressed by the remaining existence axioms. These consist of axioms asserting the existence of the pair, union and power sets, along with infinitely many further axioms of each of two kinds. First, for any condition $\Phi(x)$ there is an axiom asserting that, for any set u , *separation* from u of those elements x for which $\Phi(x)$ holds forms a set:

$$\{x \in u \mid \Phi(x)\} = \{x \mid x \in u \text{ and } \Phi(x)\}$$

Second, for any binary condition Ψ there is an axiom asserting that if for every x there is a unique $y = \psi(x)$ such that $\Psi(x, y)$ holds, then for any set u , *replacement* of each element x of u by $\psi(x)$ forms a set:

$$\{\psi(x) \mid x \in u\} = \{y \mid \Psi(x, y) \text{ for some } x \in u\}$$

(In a rigorous formalization, the vague, intuitive notion of ‘condition’ would be explained in terms of a precise, logical notion of a ‘formula’ of a symbolic language.)

To develop mathematics on the basis of these axioms, for each system of numbers, points or other mathematical objects, a system of set-theoretic objects must be specified to serve as substitutes. Usually the system of set-theoretic substitutes is easy to specify intuitively, in terms of the relationship of the substitutes to the original objects, but it may be less easy to define formally, in terms taken only from the language of set theory. Set-theoretic counterparts to the basic operations on the original objects must also be defined in the language of set theory, and the counterparts of the basic laws for these operations deduced from the axioms of set theory. When this is done, the set-theoretic substitutes are thereafter called by the name of the mathematical objects in question, which are said to have been ‘identified with’ their set-theoretic substitutes.

No more can be done here than to describe intuitively for the natural, integral, rational, real and complex numbers the related sets with which these numbers are identified. First, each natural number $n \in \mathbb{N}$ is identified with the set of natural numbers less than n , so that $0 = \emptyset$, $1 = \{0\} = \{\emptyset\}$, $2 = \{0, 1\} = \{\emptyset, \{\emptyset\}\}$, and so on; and $\mathbb{N}$ itself with the set whose existence is asserted by the axiom of infinity (above). For other kinds of numbers, one needs the identification of the ordered pair (a, b) of a, b with the set $\{\{a\}, \{a, b\}\}$, which allows one to define

$$A \oplus B = \{(0, a) \mid a \in A\} \cup \{(1, b) \mid b \in B\}$$

$$A \otimes B = \{(a, b) \mid a \in A \text{ and } b \in B\}$$

$$B^A = \{F \subseteq A \otimes B \mid \text{for all } a \in A \text{ there is a unique } b \in B \text{ with } (a, b) \in F\}$$

$$\text{dom}(F) = \{a \mid (a, b) \in F \text{ for some } b\}$$

$$\text{ran}(F) = \{b \mid (a, b) \in F \text{ for some } a\}$$

$$F \upharpoonright C = \{(a, b) \in F \mid a \in C\}$$

$$F[C] = \{b \mid (a, b) \in F \text{ for some } a \in C\}$$

$$F^{-1}[C] = \{a \mid (a, b) \in F \text{ for some } b \in C\}$$

and to define a relation on A or a function from A to B to be a subset of $A \otimes B$ or an element of B^A , respectively, so that, intuitively, a relation R or function F is identified with the set of pairs (a, b) such that aRb or $F(a) = b$. (Then the notions $\text{dom}(F)$, $\text{ran}(F)$, $F \upharpoonright C$, $F[C]$ and $F^{-1}[C]$ agree with certain customary notions from

mathematical analysis, where they are called the ‘domain’ (or set of arguments), ‘range’ (or set of values), ‘restriction’, ‘image’ and ‘pre-image’.)

Now an integer w can be identified with the set of pairs of naturals whose difference is w ; a rational x with the set of pairs of integers whose quotient is x ; a real y with the set of rationals $x < y$ or, alternatively, with the set of functions f from naturals to rationals such that $f(x)$ for larger and larger x gives a better and better approximation to y ; and a complex z with the ordered pair of its real and imaginary parts (each a real number). Ultimately any hypothesis that can be stated in conventional mathematical language can be stated in the language of set theory, and any theorem provable by conventional mathematical methods can be proved from the axioms ZF, with AC, or one of its many known implications, such as ‘dependent choice’ (DC), being needed in certain branches of higher mathematics. (DC asserts that if R is a relation on A such that for every $a \in A$ there exists at least one $b \in A$ with aRb , then there exists a function F from $\mathbb{N}$ to A with $F(n)RF(n+1)$ for all $n \in \mathbb{N}$.)

2 Ordinal numbers and ordinal arithmetic

One equivalent of AC that is widely used in mathematics is called ‘Zorn’s lemma’. A ‘partial order’ on a set A is a relation R such that for all $a, b, c \in A$ one has

aRa (reflexivity)

If aRb and bRa then $a = b$ (antisymmetry)

If aRb and bRc then aRc (transitivity).

R is an ‘order’ (also called a ‘total order’ or ‘linear order’) on a given $C \subseteq A$ if for any $a, b \in C$ either aRb or bRa (dichotomy). If R is an order on (all of) A , one often writes ‘ $\leq$ ’ (‘less than or equal to’) for R , and uses the common notations and terminology ‘ $<$ ’, ‘ $\geq$ ’, ‘ $>$ ’ and so on in the usual way. Even if R is only a partial order on A , one can use such notions as that of an ‘upper bound’ d for $C \subseteq A$ (cRd for all $c \in C$), but one must be careful to distinguish the notion of a ‘maximum’ element ($d \in C$ such that cRd for all $c \in C$) from that of a ‘maximal’ element ($d \in C$ such that, for any $c \in C$, dRc implies $c = d$), and similarly for ‘minimum’ and ‘minimal’.

Zorn’s lemma states that if R is a partial order on A and every $C \subseteq A$ on which R is a total order has an upper bound, then A has a maximal element.

The equivalent of AC most widely used in set theory is the ‘well-ordering’ principle, which asserts that every set can be well-ordered. A ‘well-order’ on a set A is a total order which also has the property that every non-empty subset of A has a minimum element (groundedness).

Cantor introduced two kinds of transfinite numbers: cardinal and ordinal. The cardinal (number) of a set is concerned only with the ‘size’ of the set; an ordinal also with the relationship of its elements. To define an ordinal, we must first define isomorphism. A function $F : A \rightarrow B$ is a ‘bijection’ (that is, a one-one, onto mapping) if for every $b \in B$ there is exactly one $a \in A$ with $F(a) = b$. Given (total) orders R, S on sets A, B , a function $F : A \rightarrow B$ ‘preserves order’ if for any $a, a' \in A$ one has aRa' if and only if $F(a)SF(a')$. An ‘isomorphism’ is a bijection that preserves order. The ‘order type’ of an order R is written $|R|$. $|R| = |S|$ if there is an isomorphism from R to S , and $|R| \leq |S|$ if there is an isomorphism from R either to S or to an initial sub-order of S , $S_b = S \upharpoonright B_b$. An ‘initial sub-order’ is the order on the ‘initial subset’ $B_b = \{b' \in B | b'Sb \text{ and } b' \neq b\}$, for some $b \in B$. An ‘ordinal’ is the order type of a well-order.

If ρ is the ordinal of a well-order on a set of cardinality α , then ρ will also be said to be of cardinality α . By the well-ordering principle, every cardinal is that of some ordinal; and it follows from the definitions that if ordinals ρ, Σ are of cardinality α, β and $\rho \leq \Sigma$, then $\alpha \leq \beta$. (To avoid confusion arising from the traditional use of the same notations for both cardinal and ordinal numbers, we will use the letters α, β, γ for cardinals, and ρ, Σ, τ for ordinals.)

Cantor proved that the order on the ordinals (defined above) is a well-order. Hence there is a least or zero ordinal. For every ordinal ρ there exists a greater ordinal and hence, again by the definition of a well-order, a *least* greater ordinal, called its successor, $\rho^\#$. For every set T of ordinals there is an ordinal greater than or equal to each of its elements, hence a least such ordinal, called its ‘supremum’, $\sup T$. (Note this implies that there can be no set of all ordinals.) A ‘limit’ (ordinal) is one that is neither zero nor a successor.

The ordinals $0, 1 = 0^\#, 2 = 1^\#, \dots$ less than the least limit are called ‘finite’, and the least limit $\sup\{0, 1, 2, \dots\}$, the ordinal of the natural order on the natural numbers, is called ω_0 or simply ω . It follows from the definitions that if Σ is the ordinal of a well-order S on a set B , then there is an isomorphism F from S to the order on the set $\{\rho \mid \rho < \Sigma\}$ of ordinals less than Σ (namely, $F(b) = |S_b|$ for $b \in B$). It is now customary to identify the ordinal Σ with this set of lesser ordinals (generalizing what was done in the case of finite ordinals - natural numbers - in §1). With this identification, it is easy to describe explicitly $\rho^\#$ and $\sup T$: they are just $\rho \cup \{\rho\}$ and $\cup T$, respectively. Thus

$$\omega^\# = \{0, 1, 2, \dots, \omega\}$$

$$\omega^{\#\#} = \{0, 1, 2, \dots, \omega, \omega^\#\}$$

and so on; and the next limit is

$$\omega^\dagger = \{0, 1, 2, \dots, \omega, \omega^\#, \omega^{\#\#}, \dots\}.$$

Addition and multiplication for order types can be defined as follows. Given linear orders R, S on sets A, B , one defines orders $R \oplus S, R \otimes S$ on $A \oplus B, A \otimes B$ by

$$\begin{aligned} (a, b)(R \oplus S)(c, d) & \quad \text{if } (a = c = 0 \text{ and } bRd) \\ & \quad \text{or } (a = 0 \text{ and } c = 1) \\ & \quad \text{or } (a = c = 1 \text{ and } bSd) \\ (a, b)(R \otimes S)(c, d) & \quad \text{if } bSd \\ & \quad \text{and (if } b = d \text{ then } aRc). \end{aligned}$$

Intuitively, $R \oplus S$ is like a copy of A in the R order, followed by a copy of B in the S order, while $R \otimes S$ is like a copy of B but with each element replaced by an entire copy of A . Then $|R| + |S| = |R \oplus S|$ and $|R| \cdot |S| = |R \otimes S|$.

For ordinals the methods of proof by transfinite induction and definition by transfinite recursion, analogous to the ordinary methods of induction and recursion in higher arithmetic or number theory, are available. To prove by transfinite induction that $\Phi(\rho)$ holds for all ordinals ρ , one must prove, as in finite induction, that $\Phi(0)$ holds and if $\Phi(\rho)$ holds then $\Phi(\rho')$ holds, and additionally that if τ is a limit and $\Phi(\Sigma)$ holds for all $\Sigma < \tau$, then $\Phi(\tau)$ holds.

Similarly, to define by recursion $\psi(\rho)$ for all ordinals ρ , one must define (1) $\psi(0)$, (2) $\psi(\rho')$ in terms of $\psi(\rho)$, and (3) $\psi(\tau)$ in terms of $\psi(\Sigma)$ for $\Sigma < \tau$ at limits τ . Equivalent, more convenient, definitions of addition and multiplication, plus a definition of ordinal exponentiation, can be given for ordinals by transfinite recursion:

$$\begin{aligned} \rho + 0 &= \rho & \rho + (\Sigma^\#) &= (\rho + \Sigma)^\# \\ \rho \cdot 0 &= 0 & \rho \cdot (\Sigma^\#) &= (\rho \cdot \Sigma) + \rho \\ \rho^0 &= 1 & \rho^{(\Sigma^\#)} &= (\rho^\Sigma) \cdot \rho \end{aligned}$$

and at the limits:

$$\begin{aligned} \rho + \tau &= \sup\{\rho + \Sigma \mid \Sigma < \tau\} \\ \rho \cdot \tau &= \sup\{\rho \cdot \Sigma \mid \Sigma < \tau\} \\ \rho^\tau &= \sup\{\rho^\Sigma \mid \Sigma < \tau\}. \end{aligned}$$

If ρ, Σ both have cardinality α , then so do $\rho + \Sigma, \rho \cdot \Sigma$ and even (in contrast to cardinal exponentiation) ρ^Σ . Note $\rho^\# = \rho + 1$. Also $\omega^\#, \omega^{\#\#}$ and $\omega^\dagger$ are $\omega + 1, \omega + 2, \omega + \omega$.

3 Cardinal numbers and cardinal arithmetic

The cardinal number (‘size’) of a set is written $||A||$. The order ‘ $\leq$ ’ on the cardinals is defined as follows. $||A|| = ||B||$ if there is a bijection from A to B , and $||A|| \leq ||B||$ if there is a bijection from A to a subset of B . This order also has the properties of a well-order, by its connection with ordinal order. Hence there is a least or

zero cardinal. Cantor proved that for any cardinal α there is a greater cardinal and hence a least greater cardinal, its successor, α^+ (an explicit description is given below). For every set C of cardinals there is a cardinal greater than or equal to each of its elements, hence a least such cardinal, the supremum of C , $\sup C$. If $C = \{\beta \mid \beta = ||B|| \text{ for some } B \in X\}$, then $\sup C = ||\cup X||$. (Note this implies there can be no set of all cardinals.) A limit (cardinal) is one that is neither zero nor a successor.

As with ordinals, the cardinals $0, 1 = 0^\#, 2 = 1^\#, \dots$ less than the least limit are called finite, and the least limit $\sup\{0, 1, 2, \dots\}$, the cardinal of the natural numbers, is called $\aleph_0$ ('aleph-nought' or 'aleph-null'). Sets of cardinality $\aleph_0$ are called 'countable'. $\aleph_0^+$ is called $\aleph_1$, and the least ordinal of cardinality $\aleph_1$ is called ω_1 or Ω . $\aleph_1^+, \aleph_1^{++}, \dots$ are called $\aleph_2, \aleph_3, \dots$, and the next limit cardinal $\sup\{\aleph_0, \aleph_1, \aleph_2, \dots\}$ is called $\aleph_\omega$, while the least ordinals of these cardinals are called $\omega_2, \omega_3, \dots, \omega_\omega$. There is only one ordinal of each finite cardinality (since there is an isomorphism between any two orders on a finite set). Not so for infinite cardinals, since the ordinals $\omega^\#, \omega^{\#\#}, \dots, \omega^\dagger$ can also be described as the ordinals of certain unnatural well-orders on the natural numbers, namely

$1, 2, 3, \dots, 0$

$2, 3, 4, \dots, 0, 1$

$1, 3, 5, \dots, 2, 4, 6, \dots, \dots$

and hence all have cardinality $\aleph_0$. Indeed, Cantor gave an explicit description of α^+ by proving that it is the cardinal of the set of ordinals of cardinality α . It is now customary to identify a cardinal with the least ordinal of that cardinal, so that $\aleph_0 = \omega_0$, and so on.

Cardinal addition, multiplication and exponentiation are defined as follows: $||A|| + ||B|| = ||A \oplus B||$, $||A|| \cdot ||B|| = ||A \otimes B||$ and $||A||^B = ||A^B||$. These agree with the usual arithmetic notions on natural numbers, and some of the usual arithmetic laws generalize from the finite to the infinite (the associative, commutative and distributive laws of addition and multiplication, and the laws of exponents). But there are some differences. Bijections from $\mathbb{N}$ to $\mathbb{N} \oplus \mathbb{N}$ and to $\mathbb{N} \otimes \mathbb{N}$ are suggested by the following.

$(0, 0) \quad (0, 0)$

$(1, 0) \quad (0, 1)$

$(0, 1) \quad (1, 1)$

$(1, 1) \quad (1, 0)$

$(0, 2) \quad (0, 2)$

$(1, 2) \quad (1, 2)$

$(0, 3) \quad (2, 2)$

$(1, 3) \quad (2, 1)$

$(0, 4) \quad (2, 0)$

$(1, 4) \quad (0, 3)$

$\vdots \quad \vdots$

This shows that $\aleph_0 + \aleph_0 = \aleph_0 \cdot \aleph_0 = \aleph_0$. This result can be generalized in two directions to show (1) that a union of countably many countable sets is countable and (2) that

$$(\aleph_0)^n = \aleph_0 \cdot \dots \cdot \aleph_0 = ||\mathbb{N} \otimes \dots \otimes \mathbb{N}|| = \aleph_0$$

for all $n \in \mathbb{N}$, and hence (3) writing $\mathbb{N}^{<\omega}$ for the set of all finite ordered sequences (pairs, triples, ...) of natural numbers and $\aleph_0^{<\omega}$ for $\sup\{(\aleph_0)^n \mid n = 0, 1, 2, \dots\}$, one has

$$\aleph_0^{<\omega} = ||\mathbb{N}^{<\omega}|| = \aleph_0.$$

These 'abstract' results have 'concrete' counterparts for the sets $\mathbb{Z}$, $\mathbb{Q}$ and $\mathbb{A}$ of integers, rational and algebraic numbers (these last being the real numbers that are roots of polynomial equations with rational coefficients); namely,

$$||\mathbb{Z}|| = ||\mathbb{Q}|| = ||\mathbb{A}|| = \aleph_0.$$

Finally, there is the generalization that $\alpha + \alpha = \alpha \cdot \alpha = \alpha^{<\omega} = \alpha$ for all infinite α , and $\alpha + \beta = \alpha \cdot \beta = \sup\{\alpha, \beta\}$ for all infinite α, β , so cardinal addition and multiplication trivialize.

Exponentiation is not trivial. A bijection between subsets of A and functions from A to $\{0, 1\}$ is obtainable by relating a subset B with its ‘characteristic function’, χ_B .

$$\chi_B(a) = \begin{cases} 1 & \text{if } a \in B \\ 0 & \text{if } a \notin B \end{cases}$$

Thus $|\wp(A)| = 2^{|A|}$. Cantor proved that $|A| < 2^{|A|}$. (Indeed, no function $F : A \rightarrow \wp(A)$ can be a bijection, since $F(a) \neq \{x \in A | x \notin F(x)\}$ for any a .)

The cardinal c (for ‘continuum’) is defined to be $|\mathbb{R}|$, where $\mathbb{R}$ is the set of real numbers (or points in the line). By the general result $|A| \cdot |A| = |A|$, c is equal to $|\mathbb{R} \times \mathbb{R}|$ or $|\mathbb{C}|$, where $\mathbb{C}$ or $\mathbb{R} \otimes \mathbb{R}$ is the set of complex numbers (or points in the plane). It can also be shown to be equal to $|\wp(\mathbb{N})|$ or $2^{\aleph_0}$, and by the general result $|A| < 2^{|A|}$ one has $\aleph_0 < c$. A ‘concrete’ consequence of this ‘abstract’ result is that there exist transcendental (that is, non-algebraic real) numbers.

The ‘generalized continuum hypothesis’ is the conjecture that $2^\alpha = \alpha^+$ for all cardinals α . The ‘continuum hypothesis’ (CH) is the special case for $\alpha = \aleph_0$. Equivalently, CH asserts there is no intermediate cardinal α with $\aleph_0 < \alpha < c$. The most interesting alternatives to CH include the hypotheses CH’: that there is just one such intermediate; and CH#: that there are c intermediates. Where CH asserts that $c = \aleph_1$ and CH’ asserts that $c = \aleph_2$, CH# implies that $c > \aleph_1, \aleph_2, \dots, \aleph_\omega$, and more. Assuming the generalized continuum hypothesis, exponentiation simplifies.

$$\alpha^\beta = \begin{cases} \alpha & \text{if } \beta < \text{cf } \alpha \\ \alpha^+ & \text{if } \text{cf } \alpha \leq \beta < \alpha \\ \beta^+ & \text{if } \text{cf } \alpha < \beta \end{cases}$$

Otherwise, one must distinguish between a cardinal α being a limit, meaning that $\beta^+ < \alpha$ whenever $\beta < \alpha$, and α being a ‘strong limit’, meaning that $2^\beta < \alpha$ whenever $\beta < \alpha$.

The ‘cofinality’ $\text{cf } \alpha$ of a cardinal α is the least β such that there is a set B of sets with $\beta = |B|$ and $|b| < \alpha$ for all sets $b \in B$, but $|\bigcup B| = \alpha$. α is ‘regular’ or ‘singular’ according as $\text{cf } \alpha$ is equal to or less than α . $\aleph_0$ is regular (otherwise $\text{cf } \aleph_0$ would be finite, but a union of finitely many finite sets is finite), $\aleph_1$ is regular (otherwise $\text{cf } \aleph_1$ would be $\aleph_0$, but a union of countably many countable sets is countable), and generally α^+ is regular. $\aleph_\omega$ is singular, with $\text{cf } \aleph_\omega = \aleph_0$ (a union of countably many sets with cardinality $\aleph_0, \aleph_1, \aleph_2$ and so on has cardinality $\aleph_\omega$). It can at least be proved that $c \neq \aleph_\omega$, and more generally that $\text{cf } 2^\alpha > \text{cf } \alpha$.

4 Combinatorial set theory

Analogous to finite combinatorics based on the ordinary arithmetic of the natural numbers there is an extensive ‘combinatorial’ set theory or infinitary combinatorics based on the transfinite arithmetic of cardinals and ordinals. A typical cardinal result is the ‘delta system’ theorem: if A is an uncountable set of finite sets, then there is a fixed finite set b and an uncountable $B \subseteq A$ such that $b' \cap b'' = b$ for all $b', b'' \in B$.

A typical ordinal result is the ‘pressing down’ theorem. Let A be a set of ordinals. A is called ‘unbounded’ in τ if $\sup(A \cap \tau) = \tau$, or for every $\rho < \tau$ there exists $\Sigma \in A$ with $\rho < \Sigma < \tau$. Given an uncountable regular cardinal α , A is called ‘closed’ in α if $\tau \in A$ whenever $\tau < \alpha$ and A is unbounded in τ . $B \subseteq \alpha$ is called ‘stationary’ in α if for all sets A that are closed and unbounded in α , $B \cap A \neq \emptyset$. The theorem asserts that if F is a function from α to α and $\{\rho < \alpha | F(\rho) < \rho\}$ is stationary, then there is a $\Sigma < \alpha$ such that $\{\rho < \alpha | F(\rho) = \Sigma\}$ is stationary.

Also included in combinatorial set theory are results about order types other than ordinals, beginning with the theory of the order types η, θ of the usual orders on $\mathbb{Q}$ and $\mathbb{R}$. Given a linear order ‘ $\leq$ ’ on a set A , an ‘open’ interval is a set of the form $]a, b[= \{x | a < x < b\}$ for $a < b$ and a ‘closed’ interval is a set $[a, b] = \{x | a \leq x \leq b\}$, $a \leq b$. $B \subseteq A$ is ‘dense’ if it is disjoint from no open interval and ‘bounded’ if it is a subset of some closed interval. $\leq$ is ‘separable’ if there exists a countable dense subset B , and satisfies the

‘countable chain condition’ (or ‘is CCC’) if there exists no uncountable set of disjoint open intervals. (Separability implies CCC since each of a set of disjoint open intervals must contain a distinct element of the countable dense set.) ‘ $\leq$ ’ is ‘complete’ if every chain of closed intervals has a non-empty intersection. $\mathbb{Q}$, $\mathbb{R}$ are each unbounded and dense in themselves, and $\mathbb{Q}$ is dense in $\mathbb{R}$ (and countable), so $\mathbb{R}$ is separable and CCC. $\mathbb{R}$ is complete, but $\mathbb{Q}$ is not (for example, the intervals

$$\dots [1.414, 1.415] \subseteq [1.41, 1.42] \subseteq [1.4, 1.5] \subseteq [1, 2]$$

have intersection $\{\sqrt{2}\}$). Cantor proved that any unbounded, dense linear order on a countable set has order type η . Similarly, any unbounded, dense, separable, complete linear order has order type θ . The ‘Suslin hypothesis’ is that the latter result holds with CCC in place of separability.

5 Point sets

The oldest part of set theory and the part most closely connected with the rest of mathematics (analysis, topology) is the theory of sets of points in the line, plane and so on (identified with $\mathbb{R}$, $\mathbb{R} \times \mathbb{R}$, $\mathbb{R}^n$, ...). Modern analysis and topology consider more complicated functions and sets than traditional calculus and geometry. A function $F: \mathbb{R} \rightarrow \mathbb{R}$ is ‘continuous’ if $F^{-1}[U]$ is open for every open $U \subseteq \mathbb{R}$ where an ‘open’ set is a union of open intervals and a ‘closed’ set is the complement of an open set. The Cantor ‘discontinuum’, the complement in the unit interval $[0, 1]$ of the union of the open intervals $]\frac{1}{3}, \frac{2}{3}[$, $]\frac{1}{9}, \frac{8}{9}[$, $]\frac{7}{27}, \frac{8}{27}[$, $]\frac{1}{27}, \frac{2}{27}[$, $]\frac{7}{27}, \frac{8}{27}[$, $]\frac{10}{27}, \frac{20}{27}[$ and so on, is a famous example of a closed set (see [Continuum hypothesis §2](#)) more complicated than any considered in traditional mathematics. Sets obtainable from open sets by iterated application of complementation and countable union are called ‘Borel’ sets. They occur ubiquitously in modern mathematics (as in the rigorous formulation of probability theory), while sets beyond Borel occur sporadically.

An ‘analytic’ set is the image of a Borel set under a continuous function, a ‘co-analytic’ set is the complement of an analytic set, a ‘pro-co-analytic’ (PCA) set is the image of any set under a continuous function, and a ‘co-pro-co-analytic’ (CPCA) set is the complement of a pro-co-analytic set. Sets obtainable from Borel sets by iterated application of continuous image and complementation are called ‘projective’. Projective sets are rather special compared with arbitrary sets. It can be shown that there are only c projective sets, whereas there are 2^c subsets of $\mathbb{R}$. The study of such special classifications of sets is called ‘descriptive’ set theory. In one notation, open sets are denoted Σ_1^0 , closed sets, Π_1^0 , unions of countably many closed sets, Σ_2^0 and their complements, Π_2^0 , ... Analytic, co-analytic, PCA and CPCA sets are denoted, respectively, Σ_1^1 , Π_1^1 , Σ_2^1 and Π_2^1 , ... sets.

Central to attempts to define integration for complicated functions, and length, area and so on for complicated linear, planar, ... sets, is the notion of a (complete probability) ‘measure’ on a set or ‘space’ X . Such a measure is a function μ from the set of ‘(μ -)measurable’ subsets of X to the unit interval $J = [0, 1]$, $\mu(A)$ being called the ‘(μ -)measure of $A \subseteq X$. μ has the following properties: (1) ‘normality’ - the whole space has measure one; (2) ‘non-triviality’ - any one-point set has measure zero; (3) ‘completeness’ - any subset of a set of measure zero has measure zero; (4) ‘complementarity’ - the complement of a measurable set is measurable; and (5) ‘additivity’ - the union of a countable set of measurable sets is measurable and, if the sets are disjoint, the measure of their union is the sum of their measures. The analogue of additivity with ‘countable set’ replaced by ‘set of cardinality less than that of the whole space’ may be called (5’) ‘superadditivity’. If we restrict ourselves to the unit interval, the ‘translation’ of $A \subseteq J$ by $t \in J$ is defined to be $\{a \# t | a \in A\}$, where $x \# y$ is defined to be whichever of $x + y$, $x + y - 1$ is an element of J , and one may consider the further property (6) ‘invariance’ - a translation of a measurable set is measurable, and its measure is the same. The key positive result is the existence of a unique measure on J with properties (1)-(6), called ‘Lebesgue measure’. (Analogues can also be defined on $\mathbb{R}$, $\mathbb{R} \times \mathbb{R}$ and so on.) For Lebesgue measure, $\mu(A)$ intuitively represents the probability that a point randomly chosen from J will be an element of A . One respect in which the discontinuum is ‘pathological’ is that it is large in the sense of having cardinality c but small in the sense of having Lebesgue measure 0. (Like some other positive results, the existence and uniqueness of Lebesgue measure require for their proof only DC (see §1 above), not full AC.)

To mention a negative result, the hypothesis that all sets are Lebesgue measurable (AM) fails. Let S be a choice set for $\{y \in J | (y - x) \in \mathbb{Q}\}$. If S were measurable, then using invariance it would follow that the disjoint sets of form $\{y \in J | y \# p \in S\}$ for $p \in \mathbb{Q}$ whose union is all of J , would all have the same measure ε . But then using additivity, $\mu(J)$ would have to be $\varepsilon + \varepsilon + \varepsilon + \dots$, which is 0 or ∞ according as $\varepsilon = 0$ or > 0 . The failure of

AM leads to consideration of some alternative hypotheses, notably the hypothesis (known to imply $\sim\text{CH}$ and $\sim\text{CH}^\#$) of the existence of a measure with properties (1)-(5) for which all sets are measurable, or the hypothesis (known to imply $\text{CH}^\#$) of the existence of a measure with properties (1)-(4) and superadditivity for which all sets are measurable. A more positive result using AC may also be mentioned. Given $Q \subseteq \mathbb{R}$ and a continuous F from $\mathbb{R}$ to $\mathbb{R}$ a uniformization for Q , F is a choice set P for $\{\{y \in Q \mid F(y) = F(x)\} \mid x \in Q\}$. AC immediately implies the existence of a uniformization for any Q , F .

The failure of AM also leads to consideration of some restricted hypotheses: for Γ one of the special classifications considered above, ‘ Γ -measurability’ is the hypothesis that all sets in Γ are measurable. Similarly, ‘ Γ -uniformization’ is the hypothesis that for any $Q \in \Gamma$ and any continuous F there is a uniformization $P \in \Gamma$. It can be shown that while Σ_p^1 -measurability and Π_p^1 -measurability imply each other for the same p , Σ_p^1 -uniformization and Π_p^1 -uniformization cannot both hold for the same p . Positive results of descriptive set theory obtained by the end of the 1930s include Σ_1^1 - and Π_1^1 -measurability, an example of a ‘regularity’ theorem or result of the form ‘Every set in Γ has some “nice” property’, and also Π_1^1 - and Σ_1^1 -uniformization, an example of a ‘structure’ theorem or result of the form ‘For every set in Γ there is another set in Γ related to it in some “nice” way’. All attempts to prove Σ_2^1 - and Π_2^1 -measurability, and Σ_3^1 - or Π_3^1 -uniformization have failed.

6 New axioms for set theory: large cardinals and determinacy

Beginning in about 1940, many conjectures Δ (including those mentioned at the end of §5) have been proved by mathematical logicians to be ‘undecidable’ relative to ZFC, meaning that it has been shown that if ZFC is consistent, then so are both $\text{ZFC} + \Delta$ (relative consistency of Δ) and $\text{ZFC} + \sim\Delta$ (relative independence of Δ). Note that the consistency of ZFC, let alone of $\text{ZFC} + \Delta$ (or $\text{ZFC} + \sim\Delta$), cannot be proved within ZFC itself, by celebrated results of mathematical logic (see **Gödel’s theorems**), and hence cannot be proved by ordinary mathematical methods, since these can be developed within ZFC. The most one can hope to prove is the *relative* result that if a contradiction could be deduced from $\text{ZFC} + \Delta$ (or $\text{ZFC} + \sim\Delta$), then a contradiction could be deduced from ZFC. There are two basic methods for proving relative consistency. A first method was used by Kurt Gödel to prove the consistency of CH relative to ZFC (and of AC relative to ZF) (see **Constructible universe**). A second method was used by Paul Cohen to prove the independence of CH relative to ZFC (and of AC relative to ZF) (see **Forcing**). Both methods have been widely applied since.

For some hypotheses, however, there can be no hope of a proof of relative consistency, namely for any hypothesis that itself implies the consistency of ZFC. An example is the hypothesis of the existence of an ‘inaccessible’ or regular strong limit cardinal α (IC). (How IC implies the existence of a model of ZFC, and hence the consistency of ZFC, will be indicated in §7.) Even in such cases there may be heuristic or inductive evidence of consistency (failure to deduce a consequence in years of work with the hypothesis). In order to decide questions undecidable in ZFC, new axioms have from time to time been advocated. The two kinds most often advocated (which unfortunately still leave CH undecided) have been ‘large cardinal’ and ‘determinacy’ axioms. The former, which pertain to sets larger than any encountered in ordinary mathematical practice, are often advocated on the intrinsic ground that they are further expressions of intuitions only partly expressed by the ZFC axioms (to the effect that the hierarchy of sets is ‘tall’). The latter, which pertain to special classifications of sets of linear, plane, ... points, and are connected with problems in analysis and topology, are often advocated on the extrinsic ground that their consequences are ‘nice’.

Large cardinal axioms include IC and also MC, the hypothesis that there exists a measurable cardinal β . Here β is measurable if there is a superadditive measure (as defined in §5), taking only the two values 0 and 1, for which all subsets of β are measurable. It follows from non-triviality and superadditivity that any subset of cardinality $< \beta$ will have measure 0, and by superadditivity again that any union of a set of cardinality $< \beta$ of such subsets will still have measure 0, whence β is regular. It also follows that if $\alpha < \beta$ and F is a function from β to $\wp(\alpha)$, then for any $\rho < \alpha$ one of the sets

$$D^+(\rho) = \{\Sigma < \beta \mid \rho \in F(\Sigma)\}$$

$$D^-(\rho) = \{\Sigma < \beta \mid \rho \notin F(\Sigma)\}$$

will have measure 0, the other, measure 1, and by superadditivity the union D of the ones with measure 0 will have

measure 0. But for $\Sigma, \tau \notin D$ one has

$$F(\Sigma) = F(\tau) = \{\rho < \alpha \mid \mu(D^+(\rho)) = 1\}$$

so F is not a bijection, $2^\alpha = ||\wp(\alpha)|| < \beta$, and β is a strong limit. So a measurable cardinal β is inaccessible, and indeed it can even be shown that $\{\alpha < \beta \mid \alpha \text{ is an inaccessible cardinal}\}$ has cardinal β , so MC is a much stronger hypothesis than IC. Still stronger is the hypothesis (SC) that there exists a ‘supercompact’ cardinal γ . For such a γ there is for any $\delta > \gamma$ an analogue of a measure as above for the subsets of δ of cardinality γ .

Determinacy axioms pertain to certain infinite games considered in mathematical game theory. Given $A \subseteq \mathbb{R}$ there is a game for two players I, II: they alternately pick the terms in the decimal expansion of a real x , and I or II wins according as x is or is not an element of A . A ‘strategy’ for a player is, intuitively, a rule telling the player how to play as a function of the opponent’s previous moves, and is winning if by so playing the player will win the game, no matter what the opponent’s moves. A is ‘determinate’ if one or other player has a winning strategy. Π_1^1 -AD, PD and AD (to add to the notation set up in §5) are, respectively, the hypotheses that all Π_1^1 , projective or arbitrary sets are determinate. AD is refutable using AC, among other reasons because AD implies AM, while, by essentially the same proof, PD implies projective measurability, and various other regularity theorems. Similarly, PD implies projective uniformization and various other structural theorems. Π_1^1 -AD yields, correspondingly, Σ_2^1 - and Π_2^1 -measurability and Π_3^1 -uniformization.

Despite superficial differences, there are deep connections between large cardinals and determinacy. By the 1970s it was established that a large cardinal axiom somewhat weaker than MC implies a conclusion somewhat stronger than Π_1^1 -AD, with the partial converse that such a determinacy hypothesis implies the consistency of such a large cardinal hypothesis. In the 1980s an analogous relation between SC and PD was established. So the cases for large cardinal and for determinacy axioms reinforce each other. The search for further axioms to decide CH continues.

7 Models of set theory

In Gödel’s method for proving relative consistency of a conjecture Δ , one seeks a formula $\Xi(x)$ and proves that for any axiom of $ZFC + \Delta$ and hence for any finite conjunction Ψ of such axioms, Ψ^Ξ is provable in ZFC, where Ψ^Ξ is the result of replacing each quantifier ‘for all x ’ or ‘for some x ’ in Ψ by the ‘restricted’ quantifier ‘for all x such that $\Xi(x)$ ’ or ‘for some x such that $\Xi(x)$ ’.

Supposing this had been done, if a contradiction could be deduced from $ZFC + \Delta$, the deduction would use only finitely many of the axioms of $ZFC + \Delta$, and so a contradiction Θ and $\sim\Theta$ could be deduced from some finite conjunction Ψ of axioms of $ZFC + \Delta$ and, restricting quantifiers, a contradiction Θ^Ξ and $\sim\Theta^\Xi$ could be deduced from Ψ^Ξ . By exhibiting the deduction of a contradiction from Ψ^Ξ , $\sim\Psi^\Xi$ would be provable in ZFC, and hence a contradiction (both Ψ^Ξ and $\sim\Psi^\Xi$) would be deducible in ZFC. Such a Ξ (and derivatively, the ‘universe’ of all x such that $\Xi(x)$ holds) is said to constitute an ‘inner model’.

In Cohen’s method for proving relative consistency, one assumes the existence of a set X that is a model of ZFC in the sense that for any axiom of ZFC, and hence for any finite conjunction Φ of such axioms, Φ^X holds, where Φ^X is the result of replacing each quantifier in Φ by the restricted quantifier (as above). One then proves the existence of a set Y that is a model of $ZFC + \Delta$, but in such a way that for any finite conjunction Ψ of axioms of $ZFC + \Delta$, the proof that Ψ^Y holds depends only on the assumption that Φ^X holds for some finite conjunction $\Phi = \Phi_\Psi$ of axioms of ZFC. Supposing this done, one uses the fact that while the existence of a model X of all of ZFC cannot be proved in ZFC (since this would imply the consistency of ZFC), for any finite conjunction Φ of axioms of ZFC the existence of a model of Φ can be proved in ZFC (as will be indicated below). It follows that the existence of a model for any finite subset Ψ of $ZFC + \Delta$ is provable in ZFC. Then if a contradiction could be deduced from $ZFC + \Delta$, a contradiction could be deduced from some finite conjunction Ψ of axioms of $ZFC + \Delta$, and by exhibiting this deduction the non-existence of any model of Ψ could be proved in ZFC, and hence a contradiction (existence and non-existence of a model of Ψ) could be proved in ZFC.

The general theories of set models X (as in the second method) and formula models Ξ (as in the first method) are closely parallel, and only the theory of set models will be sketched here. For applications it suffices to consider only sets X that are ‘transitive’ in the sense that whenever $u \in v \in X$, then $u \in X$. Informally, if Φ^X holds, it will be said that it ‘appears’ to X that Φ holds, or that Φ holds ‘inside’ X . Central to the theory of models is the

determination of sufficient conditions for various axioms of ZFC to hold inside X . Central to this determination are considerations of absoluteness, where a formula Φ is called ‘absolute’ for X if for any $u, v, \dots \in X$, $\Phi(u, v, \dots)$ appears to X to hold if and only if it really holds. This will always be so if Φ is a ‘limited’ formula, mentioning only u, v , and their elements, elements of elements and so on, since by transitivity these will all be elements of X .

For instance, the condition ‘ u, v have different elements’ or ‘For some w , either $w \in u$ and $w \notin v$ or $w \in v$ and $w \notin u$ ’ is limited, and so $u, v \in X$ appear to X to have different elements if and only if they really do have different elements, and so the axiom of extensionality, ‘For every u, v , if $u \neq v$ then u, v have different elements’, holds inside X for any transitive X . Similarly, it can be shown that the axiom of foundation holds inside X for any transitive X . As for pairing, the condition ‘ $w = \{u, v\}$ ’ or ‘ $u \in w$ and $v \in w$ and for all $z \in w$, $z = u$ or $z = v$ ’ is limited, and so w appears to X to be the pair $\{u, v\}$ of u, v if and only if it really is that pair, and so the axiom of *pairs*, ‘For every u, v there exists a w such that $w = \{u, v\}$ ’, holds inside X if and only if $\{u, v\} \in X$ whenever $u, v \in X$, or, as is said, if and only if X is closed under the pairing operation $\{, \}$. Similarly it can be shown that the axiom of *union* holds in X if and only if X is closed under the union operation $\cup$; that the axiom of infinity holds in X if and only if $\omega \in X$; and that the axiom of choice holds in X if and only if every partition in X has a choice set in X .

As for power, u appears to X to be a subset of v if and only if it really is, but w appears to X to be the set $\wp(v)$ of all subsets u of v if and only if w really is the set $\wp(v) \cap X$ of all subsets u of v with $u \in X$. Thus the axiom of *power* appears to X to hold if and only if X is closed under the operation of forming $\wp(v) \cap X$ from v , which may be the case even if X does not contain all subsets of v and therefore is not closed under the operation of forming $\wp(v)$ from v . Similarly, a necessary and sufficient condition for separation to hold inside X is that for each condition Φ , X is closed under the operation that applies to any v forms $\{u \in v \mid \Phi^X(u)\}$. Similarly, a necessary and sufficient condition for replacement to hold inside X is that for each condition Φ such that for every $x \in X$ there is a unique $y = \psi^X(x) \in X$ such that $\Phi^X(x, y)$ holds, X is closed under the operation that applied to v forms $\{\psi^X(u) \mid u \in v\}$. X is ‘supertransitive’ if whenever $u \subseteq v \in X$, then $u \in X$. For supertransitive X , $\wp(v) \cap X = \wp(v)$ for any $v \in X$, so the necessary and sufficient condition for power to hold is just that X is closed under the power operation $\wp$. Supertransitivity is a more than sufficient condition for separation to hold. Similarly, a more than sufficient condition for replacement to hold is that whenever $u \subseteq X$, $\|u\| = \|v\|$ and $v \in X$, then $u \in X$.

One kind of model of a large part of ZFC is obtained as follows. For any x let the ‘transitivization’ $t(x)$ of x be $\cup\{x, \cup x, \cup \cup x, \dots\}$, the set of all elements, elements of elements and so on of x . For any uncountable cardinal α , let $H(\alpha)$ be $\{x \mid \alpha > \|t(x)\|\}$. It follows from the definition that $H(\alpha)$ is supertransitive, and it can be checked, using the necessary and sufficient conditions indicated above, that $H(\alpha)$ is a model of all axioms of ZFC except perhaps power and replacement. Moreover, if α is a strong limit, then $H(\alpha)$ will be closed under $\wp$ and hence a model of power; while if α is regular, then one will have $u \in H(\alpha)$ whenever $u \subseteq H(\alpha)$, $\|u\| = \|v\|$, $v \in H(\alpha)$ and $H(\alpha)$ will be a model of replacement. (Thus if α is inaccessible, $H(\alpha)$ is a model of the whole of ZFC.)

Another kind of model of a large part of ZFC is obtained as follows. The hierarchy of levels of sets, intuitions about which underlie the axioms (as in §1), can be defined by transfinite recursion.

$$V(0) = \emptyset$$

$$V(\rho + 1) = \wp(V(\rho))$$

$$\text{at limits } V(\tau) = \cup\{V(\Sigma) \mid \Sigma < \tau\}$$

That $V(\rho)$ is supertransitive can be proved by transfinite induction. Given the other axioms of ZFC, the principle of replacement can be proved equivalent to the following principle of ‘reflection’: for any formula Φ there are arbitrarily large ordinals ρ such that Φ is absolute for $V(\rho)$. In particular, for any finite conjunction Φ of axioms of ZFC there are arbitrarily large ordinals ρ such that Φ holds in $V(\rho)$.

While both the $V(\rho)$ (for $\rho > \omega$) and the $H(\alpha)$ (for $\alpha > \aleph_0$) are uncountable, it can be shown that if a ZFC or some variant set theory has a transitive model, then it has a countable transitive model, using a celebrated result of mathematical logic (see **Löwenheim-Skolem theorems and nonstandard models**). Since the uncountability of the

power set of the set of natural numbers is a theorem of set theory, the existence of a countable model for set theory is sometimes considered paradoxical, and called the ‘Skolem paradox’. There is, however, no genuine contradiction. (For the conditions ‘ $x = \wp(\omega)$ ’ and ‘ $||x|| > \aleph_0$ ’ are not absolute: the former holds inside a model X if $x = \wp(\omega) \cap X$, which may well be countable even though $\wp(\omega)$ is uncountable; the latter holds inside the model X if there is no surjection $F \in X$ from ω to x , which there may well not be even if there is such an $F \notin X$.)

See also: [Cantor’s theorem](#); [Logical and mathematical terms, glossary of](#)

JOHN P. BURGESS

References and further reading

- Barwise, J.** (ed.) (1977) *Handbook of Mathematical Logic*, Amsterdam: North Holland, 317-522. (A standard reference including surveys on a generous scale of the different branches of set theory, with attribution of results to their original authors and references to the original technical literature.)
- Benacerraf, P.** (1965) ‘What Numbers Could Not Be’, *Philosophical Review* 74: 47-73; repr. in P. Benacerraf and H. Putnam (eds) *Philosophy of Mathematics: Selected Readings*, Cambridge: Cambridge University Press, 2nd edn, 1983. (A philosophical puzzle about ‘identification’ as in §1.)
- Cantor, G.** (1895, 1897) ‘Beiträge zur Begründung der transfiniten Mengenlehre’, *Mathematische Annalen* 46: 481-512, 49: 207-46; trans. P.E.B. Jourdain (1915), *Contributions to the Founding of the Theory of Transfinite Numbers*, New York: Dover, 1955. (The original source for §§2-3, with an introduction setting the work in its historical context.)
- Halmos, P.** (1960) *Naive Set Theory*, New York: Van Nostrand. (A perennially popular elementary textbook, expounding for students of mathematics the development of mathematics on the basis of ZFC mentioned in §1.)
- Jech, T.** (1977) ‘About the Axiom of Choice’, in J. Barwise (ed.) *Handbook of Mathematical Logic*, Amsterdam: North Holland, 345-70. (Surveys all aspects of AC mentioned here, and many more besides.)
- Kunen, K.** (1977) ‘Combinatorics’, in J. Barwise (ed.) *Handbook of Mathematical Logic*, Amsterdam: North Holland, 370-402. (Surveys at length combinatorial set theory as mentioned in §4; and, briefly, large cardinals as in §6.)
- Maddy, P.** (1988) ‘Believing the Axioms’, *Journal of Symbolic Logic* 53: 481-511, 736-64. (Surveys axioms old and new as in §§1, 6; especially advances made in the 1980s.)
- Martin, D.A.** (1977) ‘Descriptive Set Theory’, in J. Barwise (ed.) *Handbook of Mathematical Logic*, Amsterdam: North Holland, 783-818. (Surveys theory of projective sets (as in §§5-6) through the 1970s.)
- Shoenfield, J.R.** (1977) ‘Axioms of Set Theory’, in J. Barwise (ed.) *Handbook of Mathematical Logic*, Amsterdam: North Holland, 321-44. (Surveys material mentioned in §§1-4, 6.)
- Zermelo, E.** (1908) ‘Untersuchungen über die Grundlagen der Mengenlehre, I’, *Mathematische Annalen* 65: 261-81; trans. S. Bauer-Mengelberg, ‘Investigations in the Foundations of Set Theory’, in J. van Heijenoort (ed.) *From Frege to Gödel: A Sourcebook in Mathematical Logic*, Cambridge, MA: Harvard University Press, 1967, 200-15. (The original source for the axiomatization of set theory, with an introduction setting the work in its historical context.)

Set theory, different systems of

To begin with we shall use the word ‘collection’ quite broadly to mean anything the identity of which is solely a matter of what its members are (including ‘sets’ and ‘classes’). Which collections exist? Two extreme positions are initially appealing. The first is to say that all do. Unfortunately this is inconsistent because of Russell’s paradox: the collection of all collections which are not members of themselves does not exist. The second is to say that none do, but to talk as if they did whenever such talk can be shown to be eliminable and therefore harmless. This is consistent but far too weak to be of much use. We need an intermediate theory.

Various theories of collections have been proposed since the start of the twentieth century. What they share is the axiom of ‘extensionality’, which asserts that any two sets which have exactly the same elements must be identical. This is just a matter of definition: objects which do not satisfy extensionality are not collections. Beyond extensionality, theories differ. The most popular among mathematicians is Zermelo-Fraenkel set theory (ZF). A common alternative is von Neumann-Bernays-Gödel class theory (NBG), which allows for the same sets but also has proper classes, that is, collections whose members are sets but which are not themselves sets (such as the class of all sets or the class of all ordinals).

Two general principles have been used to motivate the axioms of ZF and its relatives. The first is the iterative conception, according to which sets occur cumulatively in layers, each containing all the members and subsets of all previous layers. The second is the doctrine of limitation of size, according to which the ‘paradoxical sets’ (that is, the proper classes of NBG) fail to be sets because they are in some sense too big. Neither principle is altogether satisfactory as a justification for the whole of ZF: for example, the replacement schema is motivated only by limitation of size; and ‘foundation’ is motivated only by the iterative conception.

Among the other systems of set theory to have been proposed, the one that has received widespread attention is Quine’s NF (from the title of an article, ‘[New Foundations for Mathematical Logic](#)’), which seeks to avoid paradox by means of a syntactic restriction but which has not been provided with an intuitive justification on the basis of any conception of set. It is known that if NF is consistent then ZF is consistent, but the converse result has still not been proved.

1 Virtual collections

Sometimes when we speak of collections it is just a manner of speaking and no more: if we say that Sarah belongs to the collection of all linguists, we might just as well say straight away that she is a linguist; nothing seems to be lost in the translation. This way of regarding collections is quite old: it is how Peano conceived of them in the 1880s. If we already have before us some formal theory which does not have collection-talk in it, then we can add it straightforwardly. If $\Phi(x)$ is a formula of the theory in question, we introduce a new term $\{x: \Phi(x)\}$ (called a ‘collection term’) into the language. Collection terms are subject to the following two rules:

- (1) If $\Phi(x)$ is a formula then $x \in \{y: \Phi(y)\}$ is an abbreviation for $\Phi(x)$.
- (2) If $\Phi(x)$ and $\Psi(x)$ are both formulas of the theory then $\{x: \Phi(x)\} = \{x: \Psi(x)\}$ is an abbreviation for $\forall x(\Phi(x) \leftrightarrow \Psi(x))$.

We permit the use of collection terms only in contexts from which these two rules allow them to be eliminated. Collections used harmlessly in this way have been called ‘virtual’ by Quine and we shall follow this terminology here. Let us introduce the convention that lower-case Greek letters stand schematically for collection terms. So the second re-writing rule allows us to assert, for instance, the following *extensionality schema*:

$$\forall z(z \in \alpha \leftrightarrow z \in \beta) \rightarrow \alpha = \beta.$$

By this schematic device we generate a considerable amount of the naïve theory of collections. We can, for instance, introduce the usual Boolean operations as follows:

$$\alpha \cap \beta = \{x: x \in \alpha \wedge x \in \beta\}$$

$$\alpha \cup \beta = \{x: x \in \alpha \vee x \in \beta\}$$

$$\emptyset = \{x: x \neq x\}$$

$$\mathbf{V} = \{x: x = x\}$$

$$\{x\} = \{y: y = x\}$$

$$\{x, y\} = \{x\} \cup \{y\}$$

and so on. The theory thus generated is elegant enough in itself, but it is difficult to be satisfied with it for long. (We have to remember, for instance, that in this theory expressions such as $\{x, y\}$ are merely notational devices and do not, despite the appearance to the contrary that their ‘explicit’ definitions might suggest, have any meaning on their own.) Collection terms stand in sentences in some of the places where the grammar leads us to expect proper names; schematic letters stand where we might expect variables. There is a drive to objectify at work here; to suppose that collection terms can stand in *all* the places where names can occur. It is of course this step which, if carelessly done, leads straight to contradiction. How to do it without contradiction will be the subject of the rest of this entry.

2 Real collections

In the virtual theory just discussed the relation of membership is defined contextually. In the real theory with collections regarded as objects in their own right membership is taken by almost all authors as primitive. (Lewis (1991) is a rare exception.) Formally, this means that the language in which the theory is couched contains, in addition to the usual logical constants, a binary predicate symbol ‘ $\in$ ’. It will be convenient to introduce a few notational conventions straight away. Upper-case Greek letters such as Φ and Ψ will stand for formulas. The notation $\Phi(x)$ stands for a formula in which the variable x occurs free: it does not indicate that no other variable occurs free in the formula. If Φ is a formula, we write $\Phi^{(a)}$ for the ‘relativization’ of Φ to a , that is, the formula obtained from Φ by replacing each quantifier $\forall x$ or $\exists x$ by $(\forall x \in a)$ or $(\exists x \in a)$, as the case may be. We write $x \subseteq y$ as an abbreviation for $\forall z(z \in x \rightarrow z \in y)$. We introduce ordered pairs by means of the device

$$(x, y) = \{\{x\}, \{x, y\}\}.$$

In the virtual theory, the question of whether there are any objects which are not collections is scarcely an issue: since there are not really any collections, if there is nothing else then there is nothing at all; the theory therefore collapses into vacuity. In any non-trivial real theory things are different: there are plenty of collections to talk about, whether or not there are any non-collections. Since this issue arises uniformly for all the real theories we shall be discussing, it is best to address it now. The objects of a theory which are not collections are called ‘atoms’, ‘individuals’ or (even in some non-German texts) ‘*Urelemente*’. If we are to allow for their presence, we need a primitive unary predicate $\text{atom}(x)$ to represent that x is an atom. We then introduce as an axiom the *axiom of extensionality (applied form)*:

$$\forall x \forall y (\text{atom}(x) \vee \text{atom}(y) \vee \forall z (z \in x \leftrightarrow z \in y) \rightarrow x = y).$$

This asserts no more than that objects which are not atoms are collections, since (as we have already observed) it is part of what we mean by a collection that it is completely determined by its members. If the atoms are to play any role in the theory at all, we shall also need the *collection axiom*, which states that the atoms form a collection:

$$\exists y \forall x (x \in y \leftrightarrow \text{atom}(x)).$$

If, on the other hand, it is pure collections that interest us, then the beginnings of the theory are a little more concise. We do not need the predicate ‘ $\text{atom}(x)$ ’ nor the collection axiom, and we can use the simpler *axiom of extensionality (pure form)*:

$$\forall x \forall y (\forall z (z \in x \leftrightarrow z \in y) \rightarrow x = y).$$

Throughout what follows we shall state theories in their pure form and adopt the convention that if T is a pure theory (that is, one which includes the pure form of the axiom of extensionality) then TU is the corresponding applied theory (that is, the one obtained from T by deleting the pure form of extensionality and substituting the

applied form together with the collection axiom). In the virtual theory we had collection terms defined contextually, provided that they occurred in certain positions (on the right of the ‘ $\in$ ’ symbol or on either side of the ‘ $=$ ’ symbol). Now that we are regarding collections as real we cannot be so indiscriminate. Let us say that a formula $\Phi(x)$ is ‘collectivizing’ in x for a theory T if

$$T \vdash \exists y \forall x (x \in y \leftrightarrow \Phi(x)).$$

Suppose now that T includes extensionality among its axioms. If $\Phi(x)$ is collectivizing in x for T then, by extensionality, we have uniqueness:

$$T \vdash \exists! y \forall x (x \in y \leftrightarrow \Phi(x)).$$

It is therefore legitimate to introduce a term $\{y: \Phi(y)\}$ with the property that

$$\forall x (x \in \{y: \Phi(y)\} \leftrightarrow \Phi(x)).$$

We thereby retrieve the virtual theory of §1 but limited to collectivizing formulas. What remains is to specify which formulas are collectivizing; it is this that is dealt with differently in the various competing theories.

3 Zermelo-Fraenkel set theory

In 1908 Ernst **Zermelo** published the first axiomatization of set theory. His choice of axioms seems to have been driven largely by his desire to ground a proof of the well-ordering theorem (see **Axiom of choice §2**; **Set theory §2**), which he published in the same year. Zermelo’s axioms were those of ZCU^- (that is, the applied form of Z^- plus the axiom of choice) except that the axiom of infinity was replaced by Zermelo’s own form (see below).

Although Zermelo’s system is strong enough for many mathematical purposes, it is weaker than ZCU^- in one important respect: in it, it is impossible to develop a satisfactory theory of ordinal numbers (see **Set theory §2**). During the 1920s this problem was solved in two ways: ingenious methods were found, principally by Kuratowski (1921), of achieving without ordinals what is most naturally done with them; and Zermelo’s system was strengthened, principally by Fraenkel and Skolem, so as to permit the theory of ordinals to be embedded in it. Zermelo-Fraenkel set theory (ZF) was the result.

We start by listing the axioms of a fragment of Zermelo-Fraenkel set theory called Z^- .

Axiom of extensionality. As before.

Axiom schema of separation. If $\Phi(x)$ is a formula in which y does not occur free, then this is an axiom:

$$\forall z \exists y \forall x (x \in y \leftrightarrow x \in z \wedge \Phi(x)).$$

Empty set axiom: $\exists x \forall y \ y \notin x$.

Pairing axiom: $\forall x \forall y \exists z (x \in z \wedge y \in z)$.

Union set axiom: $\forall a \exists x \forall y \forall z ((z \in y \wedge y \in a) \rightarrow z \in x)$.

Power set axiom: $\forall x \exists y \forall z (z \subseteq x \rightarrow z \in y)$.

Axiom of infinity: $\exists x (\emptyset \in x \wedge \forall y \forall z ((y \in x \wedge z \in x) \rightarrow y \cup \{z\} \in x)$.

To get Z we add:

Axiom of foundation: $\forall x (x \neq \emptyset \rightarrow (\exists y \in x) x \cap y = \emptyset)$.

(See **Set theory §1** for an alternative formulation.) To get ZF we add:

Axiom schema of replacement. If $\Phi(x, y)$ is a formula in which z does not occur free then this is an axiom:

$$\forall a ((\forall x \in a) \exists! y \Phi(x, y) \rightarrow \exists z (\forall x \in a) (\forall y \in z) \Phi(x, y)).$$

To get ZFC we add:

Axiom of choice: $\forall a \exists b (\forall x \in a) (x \neq \emptyset \rightarrow \exists! y (y \in x \cap b))$.

Many other variants are possible. Below are some examples.

(1) In the presence of the replacement schema the exact form of the axiom of infinity is not critical: any axiom asserting the existence of a set which can be proved to be infinite will do. Two popular choices are the following.
Axiom of infinity (weak form):

$$\exists x(\emptyset \in x \wedge \forall y(y \in x \rightarrow y \cup \{y\} \in x).$$

Axiom of infinity (Zermelo's form):

$$\exists x(\emptyset \in x \wedge \forall y(y \in x \rightarrow \{y\} \in x).$$

(2) In ZF^- (that is, ZF without foundation) and stronger theories the pairing axiom is redundant.

(3) The separation and replacement schemata and the union set axiom can be amalgamated into the following single schema.

Axiom schema of selection and union. If $\Phi(x, y)$ is a formula in which a and b do not occur free, then this is an axiom:

$$\begin{aligned} \forall y \exists a \forall x (\Phi(x, y) \rightarrow x \in a) \rightarrow \\ \forall b \exists z \forall x (x \in z \leftrightarrow (\exists y \in b) \Phi(x, y)). \end{aligned}$$

4 Von Neumann-Bernays-Gödel class theory

Zermelo-Fraenkel set theory is adequate for almost all mathematical purposes. However, its restrictions can sometimes be inconvenient. Mathematicians who use ZF and its variants are therefore accustomed to using the device of virtual sets (see §1 above), except that in order to distinguish virtual sets from the real thing they are called virtual *classes*. If we stop there, we have a good compromise: the technical simplicity of ZF plus the convenience of virtual classes when we need them. However, the urge to treat virtual objects as real is strong, and there is a popular alternative, von Neumann-Bernays-Gödel class theory (NBG), which allows us to say the same things but with classes treated as real.

Before we state the axioms of NBG we need to express what it is for a relation to be ‘functional’: we do this by writing $\text{func}(r)$ as an abbreviation for the formula $\forall x \forall y \forall z ((x, y) \in r \wedge (x, z) \in r) \rightarrow y = z$.

Axiom of extensionality: As before.

Axiom of classes: $\exists x \forall y ((\exists z y \in z) \rightarrow y \in x)$.

Empty set axiom: $\exists x (\forall y y \notin x \wedge \exists z x \in z)$.

Pairing axiom: $(\forall x \in \mathbf{V})(\forall y \in \mathbf{V})(\exists z \in \mathbf{V}) \forall u (u \in z \leftrightarrow u = x \vee u = y)$.

Axiom schema of class existence. If Φ is any formula with free variables $v_1, v_2, \dots, v_n$ then this is an axiom:

$$\begin{aligned} (\forall v_1, v_2, \dots, v_n \in \mathbf{V}) \\ \exists z \forall x (x \in z \leftrightarrow (x \in \mathbf{V} \wedge \Phi^{\mathbf{V}})). \end{aligned}$$

Union set axiom: $(\forall x \in \mathbf{V})(\exists y \in \mathbf{V}) \forall u (u \in y \leftrightarrow \exists v (u \in v \wedge v \in x))$.

Power set axiom: $(\forall x \in \mathbf{V})(\exists y \in \mathbf{V}) \forall u (u \in y \leftrightarrow u \subseteq x)$.

Axiom of infinity: $(\exists x \in \mathbf{V})(\emptyset \in x \wedge (\forall u \in x)(\forall v \in x)(u \cup \{v\} \in x))$.

Axiom of replacement: $(\forall x \in \mathbf{V}) \forall r (\text{func}(r) \rightarrow (\exists y \in \mathbf{V}) \forall u (u \in y \leftrightarrow (\exists v \in x)(v, u) \in r))$.

Axiom of foundation: As before.

Most of these axioms are unexciting transcriptions of the axioms of ZF. However, there is one striking difference: the axiom schema of replacement has turned into a single axiom. Even more surprisingly, the axiom schema of class existence can also be replaced by a finite list of axioms, such as the following.

$$\begin{aligned}
& \exists a(\forall x, y \in \mathbf{V})((x, y) \in a \leftrightarrow x \in y) \\
& \forall a \forall b \exists c \forall x(x \in c \leftrightarrow (x \in a \wedge x \in b)) \\
& \forall a \exists b(\forall x \in \mathbf{V})(x \in b \leftrightarrow x \notin a) \\
& \forall a \exists b(\forall x \in \mathbf{V})(x \in b \leftrightarrow (\exists y \in \mathbf{V})((x, y) \in a)) \\
& \forall a \exists b(\forall x, y \in \mathbf{V})((x, y) \in b \leftrightarrow x \in a) \\
& \forall a \exists b(\forall x, y, z \in \mathbf{V})((x, y, z) \in b \leftrightarrow (y, z, x) \in a) \\
& \forall a \exists b(\forall x, y, z \in \mathbf{V})((x, y, z) \in b \leftrightarrow (x, z, y) \in a)
\end{aligned}$$

So NBG is finitely axiomatizable. This is quite striking since, by contrast, ZF cannot be finitely axiomatized (if it is consistent). And yet NBG and ZF are of equal strength, in the sense that if Φ is any formula in the language of set theory then

$\text{ZF} \vdash \Phi$ if and only if $\text{NBG} \vdash \Phi^{(\mathbf{V})}$.

And moreover, NBG is consistent if and only if ZF is consistent.

Once we have taken the step of regarding classes as objects, it is puzzling why we should constrain the class existence schema by insisting that the formula defining a class can mention only sets, not proper classes. So if we are serious about the reification of proper classes, we should replace the class existence schema with the following.

Morse-Kelley schema. If $\Phi(x)$ is any formula in which y does not occur free then this is an axiom:

$$\exists y \forall x(x \in y \leftrightarrow (x \in \mathbf{V} \wedge \Phi(x))).$$

The system thus obtained is known as Morse-Kelley class theory (MK; see Kelley 1955). It is (if consistent) a genuine extension of NBG, even with respect to sets. In other words, there are formulas Φ such that $\text{MK} \vdash \Phi^{(\mathbf{V})}$ but $\text{NBG} \not\vdash \Phi^{(\mathbf{V})}$. However, the difficulty now is to know why we should stop: if classes are sufficiently real to occur legitimately in the definitions of sets, there seems no good reason why they should not be members of still more collections. As long as we refrain from calling the new collections thus countenanced ‘sets’ or ‘classes’, there is no fresh danger of inconsistency. Theories arranged in this way are not popular, though, despite their practical convenience.

5 Other axiomatizations

So far we have said little about the motivation for the axioms. Two principles have guided traditional accounts of this question: the iterative conception; and the doctrine of limitation of size. However, it is not easy to justify all the axioms of ZF on the basis of either principle on its own: for example, the replacement schema is motivated only by limitation of size; and foundation is motivated only by the iterative conception.

Other axiomatizations have been given which are closer in spirit to these principles. Here we shall describe two. The first is Scott’s way of axiomatizing the iterative conception of set, according to which sets are divided up into ‘levels’ (see Scott 1974). The set of all levels belonging to a given level is called its ‘history’. The formal details are as follows. First define the ‘accumulation’ of a set to be the set of all the elements and subsets of its members (if this set exists):

$$\text{acc}(a) = \{x: (\exists b \in a)(x \in b \vee x \subseteq b)\}.$$

Next define the predicate ‘history(a)’ to be an abbreviation for

$$(\forall b \in a)b = \text{acc}(b \cap a).$$

Then define ‘level(b)’ to be an abbreviation for

$$\exists a(\text{history}(a) \wedge b = \text{acc}(a)).$$

Now adopt the convention that the variables V, V', V'' are restricted to range over levels, so that, for instance, $\forall V \dots$ is an abbreviation for $\forall V(\text{level}(V) \rightarrow \dots)$. Then the axioms of Z' are as follows:

Axiom of extensionality: As before.

Axiom schema of separation: As before.

Axiom of creation: $\forall x \exists V x \in V$.

Axiom of infinity: $\exists V (\forall V' \in V) (\exists V'' \in V) (V' \in V'')$.

The axioms of ZF' are as follows:

Axiom of extensionality: As before.

Axiom schema of separation: As before.

Axiom schema of reflection. If $\Phi(x)$ is any formula then this is an axiom:

$$\exists V (\forall x \in V) (\Phi(x) \rightarrow \Phi^{(V)}(x)).$$

Then Z' is equivalent to Z and ZF' is equivalent to ZF .

The second theory we shall describe is due to Ackermann (1956). It takes $\mathbf{V}$ as a primitive constant. The axioms of the system A are as follows.

Axiom of extensionality: As before.

Axiom schema of separation: As before.

Axiom of transitivity: $\forall x \forall y (x \in y \wedge y \in \mathbf{V} \rightarrow x \in \mathbf{V})$.

Axiom of subsets: $\forall x \forall y (x \subseteq y \wedge y \in \mathbf{V} \rightarrow x \in \mathbf{V})$.

Ackermann's schema. If Φ is a formula in which the free variables are $v_1, v_2, \dots, v_n$ and in which $\mathbf{V}$ does not occur, then this is an axiom:

$$(\forall v_1, v_2, \dots, v_n \in \mathbf{V}) (\forall x (\Phi(x) \rightarrow x \in \mathbf{V}) \rightarrow (\exists y \in \mathbf{V}) \forall x (x \in y \leftrightarrow \Phi(x))).$$

The motivation for A is not that of the iterative conception; rather, it is based on the idea that anything formed solely from sets is itself a set. The modified system A^* is something of a hybrid: it grafts onto A the following axiom, which is derived from the iterative conception and cannot be justified by the motivating idea of A :

Axiom of foundation: $(\forall x \in \mathbf{V}) (x \neq \emptyset \rightarrow (\exists y \in x) (y \cap x = \emptyset))$.

If Φ is a formula in which the constant $\mathbf{V}$ does not occur, then

- (1) if $Z^- \vdash \Phi$ then $A \vdash \Phi^{(\mathbf{V})}$;
- (2) $A^* \vdash \Phi^{(\mathbf{V})}$ if and only if $ZF \vdash \Phi$.

6 Anti-foundation axioms

We have already mentioned that not all the axioms of ZF can be motivated by any one account of the notion of sethood. In particular, the axiom of foundation cannot be sustained solely on the basis of the doctrine of limitation of size. But without it where do sets get their individuation? To tell whether two sets are equal we compare their members; if necessary, we compare *their* members, and so on. In the presence of the axiom of foundation, this process cannot go on indefinitely; in its absence, it can. Now if the pattern of membership exhibited by the two sets is the same, then it seems reasonable to say that they are equal: they have done all that can be expected of them, as sets, to be regarded as identical. We are therefore led to the following principle.

Axiom of extensionality (strong form). If two sets have isomorphic $\in$ -graphs, then they are equal.

(The $\in$ -graph of a set is the membership relation restricted to the set, its members, their members, and so on.) As we have already noted, this assertion is provable in ZF , but in ZF^- (ZF without foundation) we need to add it as an axiom. It ensures, for instance, that there is at most one set Ω such that $\Omega = \{\Omega\}$ (since any two such sets would have to have isomorphic $\in$ -graphs). However, it does not guarantee that there is such a set. Aczel (1988) and others have studied the consequences of adding to ZF^- not only the strong extensionality axiom just mentioned (which is essentially limitative) but also a permissive axiom generating non-well-founded sets (which

can be thought of as limit points in much the same way that irrational numbers are limits of rational numbers). The most natural of the axioms that have been studied is the following *anti-foundation axiom*:

Every graph is the $\in$ -graph of some set.

7 Quine's systems

Another, more radical strand in attempts to resolve the set-theoretic paradoxes (see [Paradoxes of set and property](#)) is represented by the various forms of the theory of types (see [Theory of types](#)). We cannot, in any of the systems we have been describing so far, collect together all the objects to form another object, but we can certainly quantify over them; in the theory of types we cannot. Objects are stratified into 'types', which are not cumulative but otherwise resemble the levels of Scott's theory; variables are labelled with superscripts to indicate the type they range over. The formula $x^m \in y^n$ is taken to be well-formed only if $n = m + 1$. Now type theory is not, according to the strict definition adopted here, a form of set theory, because it does not have the axiom of extensionality. What it has instead is a schema of axioms, one for each type:

$$\forall x^{n+1} \forall y^{n+1} (\forall z^n (z^n \in x^{n+1} \leftrightarrow z^n \in y^{n+1}) \rightarrow x^{n+1} = y^{n+1}).$$

Not to call this the axiom of extensionality may seem a pedantic cavil, particularly if the device of 'typical ambiguity' is introduced, whereby formulas are written with superscripts omitted and taken to stand schematically for any way of adding superscripts which produces a grammatical result. Thus the extensionality schema above is usually represented

$$\forall x \forall y (\forall z (z \in x \leftrightarrow z \in y) \rightarrow x = y).$$

Nevertheless, we maintain, this is not the axiom of extensionality. Typical ambiguity is not true generality. To see the difference, consider the formula $\exists y \forall x x \in y$: taken as typically ambiguous this stands schematically for such formulas as

$$\exists y^{n+1} \forall x^n x^n \in y^{n+1}$$

and is therefore true; taken as a formula in its own right it is false. The restrictions on what can be said in the theory of types have struck many writers as too severe. However, the means by which the theory avoids paradox are as secure as those of iterative set theories. Quine has proposed a system which uses type theory's method of paradox-avoidance but abandons its grammatical restrictions. Specifically, he regards formulas as genuine, not typically ambiguous, but calls a formula 'stratified' if it is possible to decorate it with superscripts so as to make it well-formed according to the theory of types. The system, which is known as NF because it was proposed in an article called '[New Foundations for Mathematical Logic](#)' (1937), has the following axioms.

Axiom of extensionality: As before.

Axiom schema of stratified comprehension. If $\Phi(x)$ is any stratified formula in which y does not occur free then this is an axiom:

$$\exists y \forall x (x \in y \leftrightarrow \Phi(x)).$$

NF is very different from ZF. For one thing NF is finitely axiomatizable. Moreover, since the formula $x = x$ is stratified, it follows at once that there is a set of all sets: if we denote it $\mathbf{V}$ then $\mathbf{V} \in \mathbf{V}$. The axiom of choice is provably false in NF. When we come to arithmetic, it is possible in a natural way to define zero and the successor operation in NF, and then to define the set of natural numbers to be the intersection of all sets containing zero and closed under successors. This works well up to a point, but gives us mathematical induction only for stratified formulas. Unfortunately some useful formulas are not stratified. For example, we cannot prove in NF, and are therefore forced to add as an axiom, the *axiom of counting*:

If n is a natural number then it has exactly n predecessors.

In an effort to overcome this inconvenience Quine has proposed a second system called ML which stands to NF in somewhat the same relation that NBG stands to ZF. The second system takes $\mathbf{V}$ as a primitive constant and has the following axioms.

Axiom of extensionality: As before.

Axiom schema of class existence. If $\Phi(x)$ is a formula in which y does not occur free then this is an axiom:

$$\exists y \forall x (x \in y \leftrightarrow y \in \mathbf{V} \wedge \Phi(x)).$$

Axiom schema of set existence. If $\Phi(x, v_1, v_2, \dots, v_n)$ is a stratified formula in which no variables other than those listed occur free, then this is an axiom:

$$\forall v_1, v_2, \dots, v_n \exists y \in \mathbf{V} \\ \forall x (x \in y \leftrightarrow \Phi^{(\mathbf{V})}(x, v_1, \dots, v_n)).$$

In contrast to NF, ML is not finitely axiomatizable if it is consistent. If Φ is any formula in which $\mathbf{V}$ does not occur then

$\text{NF} \vdash \Phi$ if and only if $\text{ML} \vdash \Phi^{(\mathbf{V})}$.

Thus ML is consistent if and only if NF is consistent. However, it is still not known whether ML and NF are consistent if ZF is consistent. If we define the natural numbers in superficially the same way as in NF, we get mathematical induction even for non-stratifiable formulas, obviating the need for an axiom of counting. However, this time we cannot prove that the class of natural numbers is a set (if ML is consistent). So now we have to add *this* as an axiom. Thus neither NF nor ML suffices without supplementation as a basis for mathematics.

See also: [Logical and mathematical terms, glossary of](#)

MICHAEL POTTER

References and further reading

- Ackermann, W.** (1956) 'Zur Axiomatik der Mengenlehre' (On the Axiomatization of Set Theory), *Mathematische Annalen* 131: 336-45. (The source for the system A mentioned in §5.)
- Aczel, P.** (1988) *Non-Well-Founded Sets*, Stanford, CA: Center for the Study of Language and Information, Stanford University. (An account of anti-foundation axioms, aimed at the mathematically aware.)
- Boolos, G.** (1989) 'Iteration Again', *Philosophical Topics* 17 (2): 5-21. (An interesting but fairly technical discussion of the iterative and limitation-of-size conceptions of set.)
- Bourbaki, N.** (1968) *Theory of Sets*, New York: Addison-Wesley. (A development of set theory in a system equivalent to ZF^- together with a strong form of the axiom of choice. The system is notable for its employment of the axiom schema of selection and union referred to in §3.)
- Forster, T.E.** (1992) *Set Theory with a Universal Set*, Oxford: Oxford University Press. (A survey of work on NF, well written but aimed at logicians.)
- Fraenkel, A.A.** (1922) 'Zu den Grundlagen der Cantor-Zermeloschen Mengenlehre' (On the Foundations of Cantor-Zermelo Set Theory), *Mathematische Annalen* 86: 230-7. (This paper is the reason Fraenkel's name is associated with the standard formulation of set theory.)
- Fraenkel, A.A., Bar-Hillel, Y. and Levy, A.** (1958) *Foundations of Set Theory*, Amsterdam: North Holland; revised 2nd edn, 1973. (Detailed discussions of most of the systems of set theory mentioned here and a wealth of historical information and references.)
- Gödel, K.** (1940) *The Consistency of the Continuum Hypothesis*, Princeton, NJ: Princeton University Press, 1994. (The definitive formulation of NBG.)
- Hallet, M.** (1984) *Cantorian Set Theory and Limitation of Size*, Oxford: Clarendon Press. (A detailed study of the early history of set theory.)
- Heijenoort, J. van** (ed.) (1967) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press. (It is possible to trace the historical development of ZF and NBG through the papers reprinted here.)
- Kelley, J.L.** (1955) *General Topology*, Princeton, NJ: Van Nostrand. (The appendix includes the first published version of MK.)
- Kuratowski, K.** (1921) 'Sur la notion d'ordre dans la théorie des ensembles' (On the Notion of Order in Set Theory), *Fundamenta Mathematicae* 2: 161-71. (Mentioned in §3.)
- Lewis, D.K.** (1991) *Parts of Classes*, Oxford: Blackwell. (An entertaining, non-technical discussion of a way of separating the hierarchical from the non-hierarchical part of set theory.)

- Morse, A.** (1965) *A Theory of Sets*, New York: Academic Press. (A treatment of the theory of classes based on a version of MK.)
- Neumann, J. von** (1925) 'Eine Axiomatisierung der Mengenlehre' (An Axiomatization of Set Theory), *Journal für die reine und angewandte Mathematik* 154: 219-40; trans. in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967. (The first formulation of an axiomatization similar to what is now called NBG.)
- Potter, M.D.** (1990) *Sets: An Introduction*, Oxford: Oxford University Press. (An account of set theory based on a Scott-type axiomatization; intended for the mathematically-minded.)
- Quine, W.V.** (1937) 'New Foundations for Mathematical Logic', *American Mathematical Monthly* 44: 70-80. (Principally of historical interest. For up-to-date information about NF consult Forster (1992).)
- Quine, W.V.** (1940) *Mathematical Logic*, Cambridge, MA: Harvard University Press; revised edn, 1951. (Very technical, but includes a wealth of information.)
- Quine, W.V.** (1963) *Set Theory and its Logic*, Cambridge, MA: Harvard University Press; revised edn, 1969. (Rather idiosyncratic but extremely clear and well written.)
- Scott, D.** (1974) 'Axiomatizing Set Theory', in *Axiomatic Set Theory*, Proceedings of the Symposium of Pure Mathematics 13 (2), Providence, RI: American Mathematical Society, 207-14. (The primary source for Scott's axiomatization of ZF; described in §5.)
- Skolem, T.** (1923) 'Einige Bemerkungen zur axiomatischen Begründung der Mengenlehre', in *Matematikerkongressen i Helsingfors den 4-7 Juli 1922, Den femte skandinaviska matematikerkongressen, Redogörelse*, Helsinki: Akademiska Bokhandeln, 217-32; trans. S. Bauer-Mengelberg, 'Some Remarks on Axiomatized Set Theory', in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967, 291-301. (The first precise formulation of the axiom schema of replacement and of Zermelo's notion of 'definite property'.)
- Zermelo, E.** (1908) 'Untersuchungen über die Grundlagen der Mengenlehre I', *Mathematische Annalen* 65: 261-81; trans. S. Bauer-Mengelberg, 'Investigations in the Foundations of Set Theory', in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967, 200-15. (The first published axiomatization of set theory.)

Syntax

Syntax (more loosely, 'grammar') is the study of the properties of expressions that distinguish them as members of different linguistic categories, and 'well-formedness', that is, the ways in which expressions belonging to these categories may be combined to form larger units. Typical syntactic categories include noun, verb and sentence. Syntactic properties have played an important role not only in the study of 'natural' languages (such as English or Urdu) but also in the study of logic and computation. For example, in symbolic logic, classes of well-formed formulas are specified without mentioning what formulas (or their parts) mean, or whether they are true or false; similarly, the operations of a computer can be fruitfully specified using only syntactic properties, a fact that has a bearing on the viability of computational theories of mind.

The study of the syntax of natural language has taken on significance for philosophy in the twentieth century, partly because of the suspicion, voiced by Russell, Wittgenstein and the logical positivists, that philosophical problems often turned on misunderstandings of syntax (or the closely related notion of 'logical form'). Moreover, an idea that has been fruitfully developed since the pioneering work of Frege is that a proper understanding of syntax offers an important basis for any understanding of semantics, since the meaning of a complex expression is compositional, that is, built up from the meanings of its parts as determined by syntax.

In the mid-twentieth century, philosophical interest in the systematic study of the syntax of natural language was heightened by Noam Chomsky's work on the nature of syntactic rules and on the innateness of mental structures specific to the acquisition (or growth) of grammatical knowledge. This work formalized traditional work on grammatical categories within an approach to the theory of computability, and also revived proposals of traditional philosophical rationalists that many twentieth-century empiricists had regarded as bankrupt. Chomskian theories of grammar have become the focus of most contemporary work on syntax.

1 The need for structure

One central aim of *semantics* is to explain how the meaning of a sentence (or any other complex expression) is a function of the meanings of its parts (see **Compositionality**), a project that presupposes an understanding of how the parts are put together, that is, an understanding of its *syntax*.

Let us call any sequence of words a 'string'. The following strings can be put together from the list of words 'Bill', 'slept', 'soundly', 'last' and 'night':

- (1) (a) Bill slept soundly last night
- (b) last night Bill slept soundly
- (c) *last soundly Bill slept night
- (d) *night Bill soundly last slept.

There is a clear difference between, on the one hand, (1a) and (1b) and, on the other, (1c) and (1d). In ordinary talk, the former are *sentences*; the latter are not. (Asterisks indicate non-sentences.) In order for a string of words to be a sentence (in this ordinary sense) the words must be put together in a certain way. Syntax is, in part, the study of the rules (or conditions) that determine the way that sentences are structured. The order of words in a string does more than determine whether or not a string is a sentence; it also plays a role in determining meaning. Compare (2a) and (2b):

- (2) (a) The dog chased the cat.
- (b) The cat chased the dog.

The fact that (2a) and (2b) differ in meaning cannot be a consequence of their containing words with different meanings, since they contain precisely the same words. The difference in meaning is attributable to a difference in *word order*. In the terminology of traditional grammar, in English we understand a noun phrase preceding a verb (in the active voice) as the subject, and the noun phrase following that verb as its direct object. Not all languages mark the subject-object distinction in this way. Thus, for example Latin marks it by inflection and this allows for the possibility of much freer word order.

Initially, the following generalizations might seem reasonable: (a) sentence meaning is the product of word meaning and word order; (b) a difference in word order results in a difference in sentence meaning. But in fact neither is accurate. Although a change in word order frequently results in a change in meaning, sentences (1a) and (1b) above show that this is not always the case, so (b) is incorrect. To see that (a) is incorrect consider

(3) Mary said Bill left voluntarily.

This admits of two distinct interpretations (or readings), which can be paraphrased as (a) ‘Mary said that Bill left and she said this voluntarily’ and (b) ‘Mary said that Bill left and that he did so voluntarily’. It is as if the word ‘voluntarily’ can be understood in connection with either the saying or the leaving. Sentence meaning, then, is the product of more than word meaning and word order; it is the product of word meaning and sentence structure, and there is more to the sentence structure than the linear order in which its words appear: the way in which the words are grouped is crucial. One way of cashing out this idea might be to say that ‘voluntarily’ can be understood as an attachment either to the sentence ‘Mary said Bill left’ or to the smaller sentence ‘Bill left’. Using square brackets to group parts of the sentence, we can represent the two different interpretations of (3) as follows:

(4) (a) [Mary said [Bill left]] voluntarily.

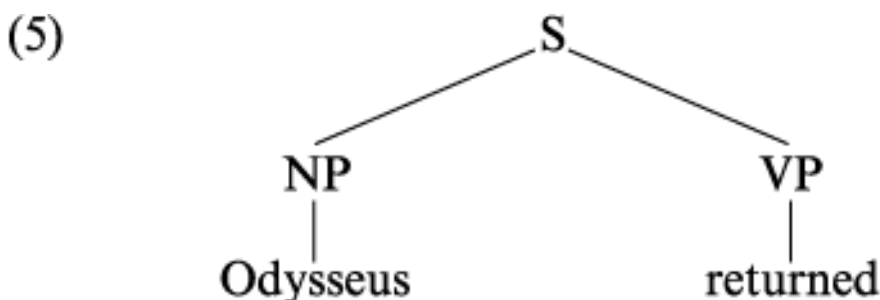
(b) Mary said [[Bill left] voluntarily].

In a sense that we can make precise, the moral of an example such as (3) is that the *hierarchical* organization (grouping) of the words in a sentence is just as important as their *linear* organization. (In terms explained in §5 below, the *scope* of ‘voluntarily’ is different in (4a) and (4b).)

2 Syntactic categories and constituent structure

The use of syntactic theory in projects such as compositional semantics requires that the theory have some systematic way of describing the syntax of sentences. The ways in which grammatical theories can describe linguistic structure are varied, but the following exposition will illustrate the most common. Since we are not here concerned with the internal structure of words (their morphology), let us say that words are the ‘ultimate constituents’ of sentences. The two parts of the sentence ‘Odysseus returned’ correspond to the traditional distinction between subject and predicate. Of course, both subject and predicate expressions can be more complex. If we replace the subject expression ‘Odysseus’ in ‘Odysseus returned’ by ‘the hero’, ‘no hero’ or ‘a hero as great as godlike Achilles’, in each case the result is another sentence. Since all of these expressions are built upon nouns, they are said to belong to the category of noun phrase (NP).

If we replace the verb ‘returned’ in the sentence ‘Odysseus returned’ by ‘loves Penelope’, or ‘shaves another customer’, in each case the result is another sentence. Since expressions in predicate position are (typically) based on verbs, they are said to belong to the linguistic category of verb phrase (VP). Using ‘S’ for ‘sentence’, we can represent the syntactic structure of the sentence ‘Odysseus returned’ using a ‘phrase structure tree’ (or ‘phrase marker’) as follows:



Alternatively, we can represent exactly the same syntactic information using a ‘labelled bracketing’:

(6) [_S [_{NP} Odysseus] [_{VP} returned]].

(5) and (6) are notational variants of one another, they are equivalent descriptions of the same structure - ‘structural descriptions’ - that is, tree notation and bracket notation are two informationally equivalent ways of specifying syntactic structure.

We can begin to characterize a set of sentences by formulating a rule to the effect that an NP followed by a VP forms a sentence, S. The standard way of doing this is to use a 'phrase structure rule' (or 'rewrite rule'):

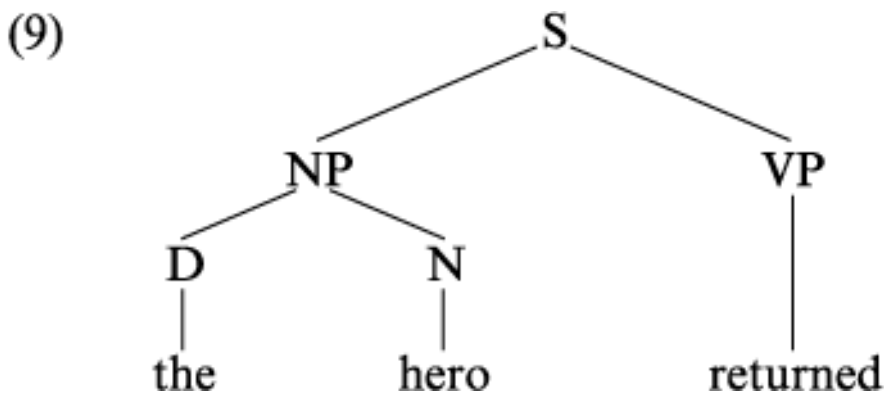
(7) $S \Rightarrow NP + VP$.

This is read as 'S goes to NP VP'. All (7) says is that a sentence, or S, may be composed of an NP followed by a VP.

Let us turn now to the internal structures of NPs and VPs. Many NPs appear to be composed of words belonging to the traditional grammatical categories of article (for example, 'the', 'a') and noun (for example, 'cat', 'man'). Rather than using the category of article, let us use the broader category of *determiner* that includes not just 'the' and 'a', but also, for example, 'every', 'some', 'no', 'neither' and 'one'. We can represent the fact that an NP may be composed of a determiner (D) and a noun (N) using the following phrase structure rule:

(8) $NP \Rightarrow D + N$.

Now we can provide a phrase structure tree for 'The hero returned' using the rules in (7) and (8):



Or we can use a labelled bracketing:

(10) $[_S [_{NP} [_D \text{the}] [_N \text{hero}]] [_{VP} \text{returned}]]$.

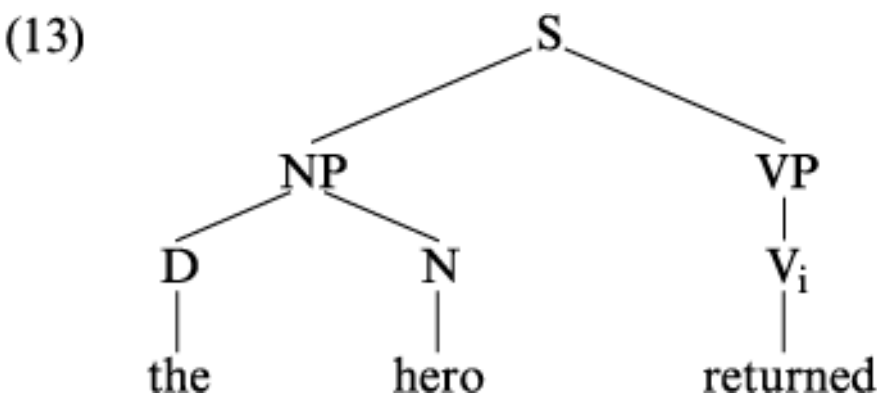
Some VPs may contain NPs as parts, for example, 'loves Penelope', and 'shaves another customer' (see above). Unlike verbs such as 'return' and 'snore', verbs like 'see' and 'respect' are *transitive*: they take NPs as direct objects. We have, then, at least two types of VPs to examine: those 'headed by' (that is, built around) intransitive verbs (V_i s), such as 'return' and 'snore', and those headed by transitive verbs (V_t s), such as 'respect' and 'like'.

Consequently, we need at least two phrase structure rules for VPs:

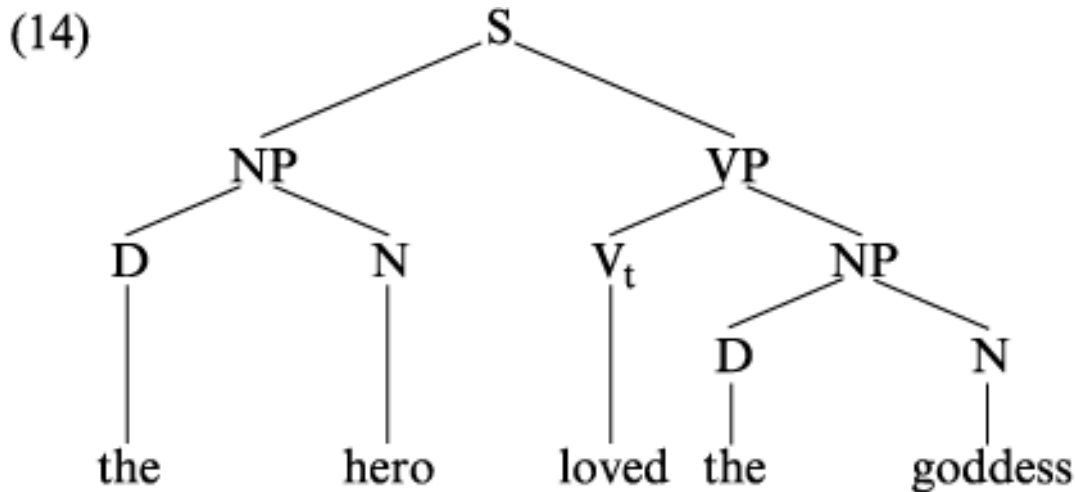
(11) $VP \Rightarrow V_i$

(12) $VP \Rightarrow V_t + NP$.

The syntax of 'The hero returned' can now be spelled out in more detail:



The syntax of 'The hero loved the goddess' is given by



3 Phrase structure grammars

Not only do philosophical applications of syntactic theory require a set vocabulary for describing linguistic structure, they require linguistic theories to specify well-formed linguistic structures in a finite way. Contemporary syntactic theory offers just such resources.

One aim of what **Chomsky** has called ‘generative grammar’ is to articulate a finitely storable theory that generates all and only the phrase markers of a given language. With the resources made available in the previous section, we can construct a generative grammar for a fragment of English. For purposes of illustration, we can do this by articulating a ‘phrase structure grammar’. Such a grammar can be viewed as a formal system consisting of two parts: a ‘lexicon’ (a list of words together with a specification of the grammatical category of each word), and a set of phrase structure rules specifying how words from these categories may be put together to form sentences. In effect, then, a grammar is a theory, and like any other theory its usefulness lies in its predictive power. For any string of English words, the theory must say whether or not that string is a sentence. The following is a simple generative grammar:

Lexicon

PN = {Fred, Mary, Bill}

D = {a, every, the}

N = {man, woman}

V_i = {left, returned}

V_t = {likes, respects}

Phrase structure rules

S ⇒ NP + VP

NP ⇒ PN (‘proper name’)

NP ⇒ D + N

VP ⇒ V_i

VP ⇒ V_t + NP

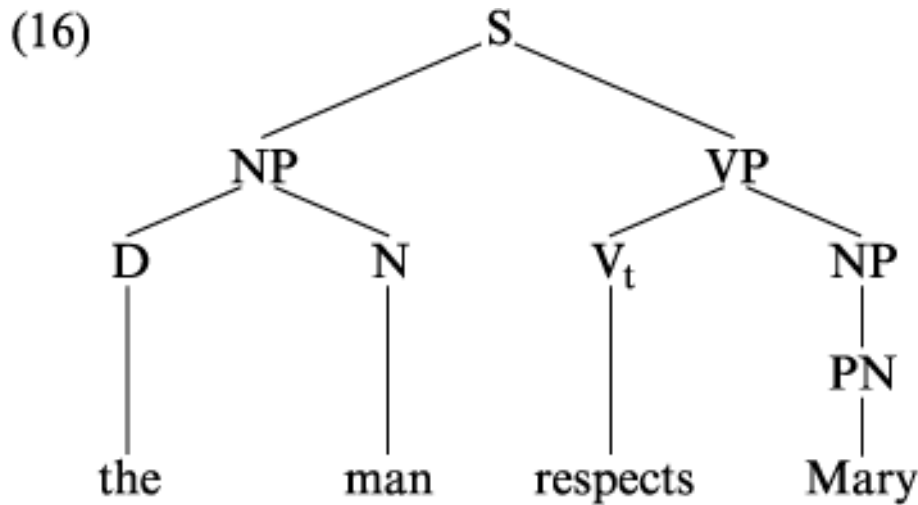
Notice that each of the phrase structure rules has the following form:

(15) $\alpha \Rightarrow \beta_1 + \dots + \beta_n$.

The important feature here is that in each rule of this type exactly one symbol α appears on the left-hand side of ‘⇒’. Such rules are called ‘context-free’ phrase structure rules, the idea behind the terminology being that they are blind to whatever symbols may appear on either side of α , that is, they are blind to the syntactic ‘context’ in which α appears. The systematic use of a set of context-free rules is a way of formalizing traditional work on

grammatical categories within an elegant approach to the theory of computability (or recursive functions); a context-free phrase structure grammar is equivalent to a categorial grammar in the sense of Ajdukiewicz (1935; see Ajdukiewicz, K. §5).

In a sense to be defined, the grammar just given generates the following phrase structure tree:



There is a specialized vocabulary in formal language theory (mathematical linguistics) for describing the parts of phrase structure trees and the relationships that obtain between the parts. We can tailor some of this vocabulary, rather informally, to suit our concerns:

(i) Each position in a tree is a *node*. A line connecting two nodes is a *branch*.

So, for example, in (1) there are twelve distinct nodes: S, NP, VP, D, N, V_t, PN, 'the', 'man', 'respects' and 'Mary'. Branches connect D and 'the', NP and D, and NP and N.

(ii) The S node at the 'top' of a tree is the *root node* of that tree. The words at the 'bottom' of a tree are its *leaf nodes* (or *terminal nodes*).

We are concerned with (a fragment of) English, so every leaf node will be a word of English.

(iii) A sequence of branches that connects two nodes is a *path*. Node β is *properly dominated* by node α if and only if the path from β back to the root node passes through α . The set of nodes properly dominated by node α is the *proper domain* of α .

Example: in (16) there is a path between D and 'the', and a path between S and D. Indeed, there is a path between any two distinct nodes. Additionally, D properly dominates 'the', NP properly dominates D, N, 'the' and 'man', and S properly dominates every node in the tree (apart from itself). (It dominates but does not properly dominate itself, just as a set is a subset but not a proper subset of itself.)

(iv) Each node in a tree corresponds to a *constituent* of the sentence in question.

So, for example, in (16) the NP node corresponds to the constituent 'the man'; and the VP node corresponds to the constituent 'respects Mary'. So whereas 'the', 'the man', 'respects' and 'respects Mary' are constituents of (16), 'man respects', 'man respects Mary' and 'the man respects' are not.

(v) If Γ is a phrase structure grammar and τ is a tree, then Γ *generates* τ if and only if

(a) the root node of τ is S;

(b) every terminal node of τ is in the lexicon for Γ ; and

(c) every step from the root node to the string of terminal nodes is licensed by one of the phrase structure rules of Γ .

(vi) A *string* of Γ is a linear sequence of words taken from the lexicon for Γ . Γ *generates a string* Σ if and only if Γ generates a tree for Σ .

(vii) The set of strings generated by a grammar Γ is the *language* L_Γ of Γ , that is, the set of *sentences* of L_Γ .

An interesting feature of (vii) is that it appropriates the English words ‘language’ and ‘sentence’ and assigns them technical meanings. Later, the theoretical use of these words will be specified precisely to satisfy theoretical demands that arise when syntax and semantics come together.

The picture just presented is deliberately simplified. Nevertheless, it is widely held that context-free grammars of this form are inadequate to describe natural languages. In addition to context-free rules of the type we have just discussed, Chomsky (1957, 1965) posited ‘transformational’ rules that operated on the trees generated by context-free phrase structure grammars to produce trees that could not be produced in any satisfactory way by phrase structure rules alone. For example, Chomsky wanted to explain the syntactic and semantic relationships between the active and passive voice and accomplished this by postulating a transformational rule that derived a phrase structure tree for a passive sentence from the tree for its corresponding active. Much work by generative linguists in the 1970s sought to impose a rigorous set of constraints on possible transformational rules. Beginning with Chomsky (1981), generative linguists began to explore alternative ways of generating sentences using methods that depart from those of traditional phrase structure grammars.

4 Recursion and sentential verbs

The grammar above was finite, as was the language that it generated (it consists of a finite number of sentences). There are, however, finite grammars that generate infinite languages. This can be seen easily enough in connection with philosophically interesting verbs such as ‘believe’, ‘know’, ‘prove’, ‘remember’ and ‘say’, which can be used in ways that preclude classifying them as either V_{is} or V_s . Consider the following sentences:

(17) Mary believes (that) Fred left.

(18) Mary said (that) Fred left voluntarily.

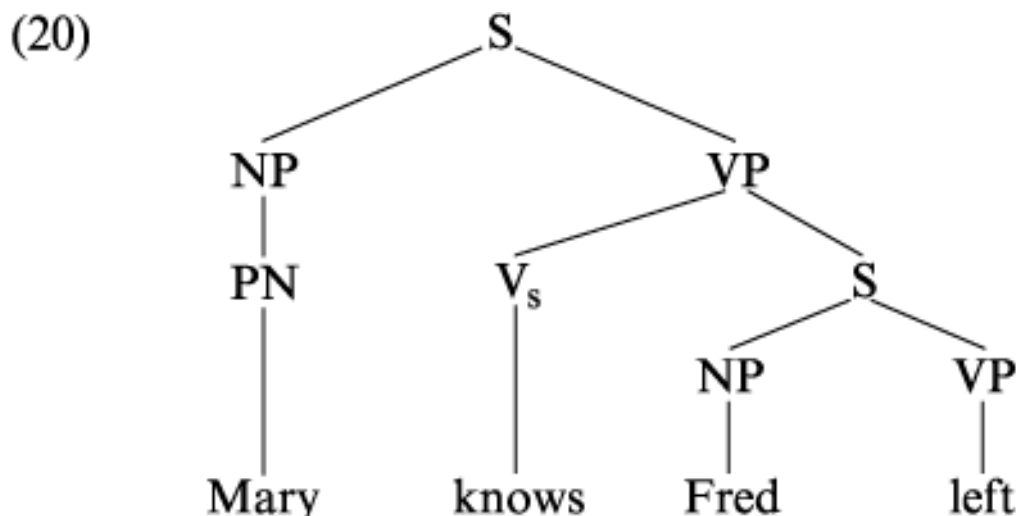
The VPs in these sentences seem to consist of a verb combined with a whole sentence. Since they seem to take whole sentences as their complements, these can be called ‘sentential verbs’ (V_s). Let us now add some V_s s to our lexicon:

$V_s = \{\text{believe, know, doubt, say, suggest, realize, remember, prove}\}$.

(Some V_s s are also known as ‘*psychological verbs*’ or ‘verbs of propositional attitude’, a label due to Russell - see **Propositional attitude statements**; §5 below.) We can now add the following phrase structure rule to produce a new grammar:

(19) $VP \Rightarrow V_s + S$.

The resulting grammar generates the following phrase marker:



As this tree diagram reveals, the rules ' $S \Rightarrow NP + VP$ ' and ' $VP \Rightarrow V_s + S$ ' taken together give our grammar the capacity to generate trees in which S nodes are properly dominated by other S nodes. More precisely, our grammar generates trees in which an S node properly dominates a VP node that properly dominates another S node that properly dominates another VP node that.... So, for example, the grammar generates trees for both of the following:

(21) Bill knows Mary thinks Fred left.

(22) Mary doubts Fred knows Bill said Mary left.

Clearly, there is no end to the list of sentences generated: if our grammar generates a sentence ϕ , it also generates every sentence of the form

$[S[NP \dots][VP[V_s \dots]\phi]]$.

Despite containing a finite set of rules, it generates an infinite set of sentences (the language generated is infinite). A structure such as (20) in which a node of category X properly dominates a distinct node also of category X is a *recursive* structure. We now say that our grammar *recursively generates* an infinite set of sentences. The recursive nature of natural language is further revealed once connectives such as 'and', 'or' and 'but' are introduced. A primary function of connectives is to join two sentences - they may also be used to connect, for example, NPs and VPs - in such a way that the result is a larger sentence, as in

(23) Bill left and Mary returned.

(24) Bill snores or Mary snores.

The following phrase structure rule can therefore be postulated:

(25) $S \Rightarrow S \text{ CONN } S$,

where CONN is the category 'connective'. This rule exploits the possibility of having three symbols on the right-hand side of ' $\Rightarrow$ ', but it is still a context-free phrase structure rule of the form given in (15). Recursive structures are also found in (for example) NPs containing relative clauses such as 'the man who loves the woman who found the child who...'.

5 Scope

There are a number of semantic issues of independent philosophical importance that can be illuminated by appealing to syntactic structure. The nature of scope in natural language is an instructive case study (see [Scope](#)). The following strings are all ambiguous:

(26) (a) everyone strives for some good

(b) if I know that p then necessarily p

(c) the number of planets is necessarily greater than seven

(d) Bill returned or Fred returned and Mary left

(e) George thought the King wasn't the King

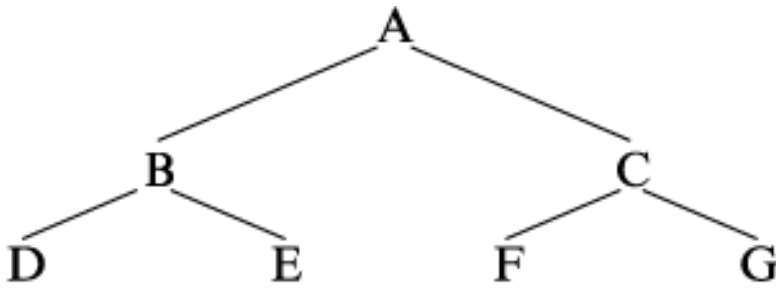
(f) Mary said Bill left voluntarily.

For example, someone uttering (26a) might be understood as saying that there is some particular good such that everyone strives for it or merely that everyone strives for some good or other. All of these examples involve ambiguities of scope that can be explicated within sophisticated syntactic theories. For present concerns, we restrict ourselves to a simple case mentioned earlier involving the scope of the adverb 'voluntarily' as it occurs in a string such as (26f), which is ambiguous between readings corresponding to the syntactic structures given earlier by (4a) and (4b). To provide a general characterization of scope in terms of phrase structure is straightforward once we introduce some new notions. First,

(viii) Node β is *immediately dominated* by node α if and only if β is properly dominated by α and there is no intervening node γ , that is, no node γ such that α properly dominates γ and γ properly dominates β .

In the following diagram, A properly dominates every other node; in addition it immediately dominates B and C,

but not D, E, F or G:



(ix) If distinct nodes β and γ are immediately dominated by node α , then α is a *branching node*.

A, B and C are branching nodes; D, E, F and G are not. We can now define ‘scope’ in a phrase structure tree:

(x) The *scope* of node α is (the constituent corresponding to) the first branching node properly dominating α . Node β is *within the scope* of node α if and only if the first branching node properly dominating α properly dominates β .

So in (20), the scope of ‘Mary’ is the entire sentence, whereas the scope of ‘Fred’ is the smaller sentence ‘Fred left’.

(xi) If node β is within the scope of node α and node α is not within the scope of node β , then α has larger scope than β (and β has smaller scope than α).

Thus ‘Mary’ has larger scope than ‘Fred’ in (20). The scope of ‘knows’ in this phrase marker is the VP ‘knows Fred left’. Thus the name ‘Fred’ occurs within the scope of ‘knows’, but the name ‘Mary’ does not. This is widely held to have important logical consequences. If Mary is Elizabeth - that is, if ‘Mary’ and ‘Elizabeth’ are two names of the same person - then the truth of (21) guarantees the truth of

(27) Elizabeth thinks Fred left.

Now suppose that Fred is Bert. Does the truth of (21) guarantee the truth of (28)?

(28) Mary thinks Bert left.

Following Frege, many philosophers believe it does not. (Mary may not know that Fred is Bert.) If this is correct, then we can describe the situation syntactically: the fact that ‘Fred’ appears within the scope of a ‘psychological’ verb (in this case, ‘think’) means there is no guarantee that ‘Fred’ can be replaced by a coreferring name to produce a sentence with the same truth-value as the original. A name occupying such a position is said to occur in a context that is ‘referentially opaque’ (see [Propositional attitude statements §§1-2](#)).

6 Ambiguities of scope

We turn now to an example involving what is usually called a ‘structural ambiguity’ or an ‘ambiguity of scope’. The latter label (due to Russell) is more dominant in philosophy, logic, and mathematics. As Russell pointed out, much bad philosophy has resulted from inattention to scope ambiguities. Indeed, uncovering and avoiding them are skills all philosophers must acquire. Davidson has emphasized the importance of an understanding of the semantics of adverbs to a number of philosophical questions involving actions and events. Earlier, the fact that the following string is ambiguous was used to call attention to the grouping of words within a sentence:

(3) Mary said Bill left voluntarily.

It was suggested that the ambiguity in question was structural in that the words can legitimately be grouped in two different ways:

(4) (a) [Mary said [Bill left]] voluntarily.

(b) Mary said [[Bill left] voluntarily].

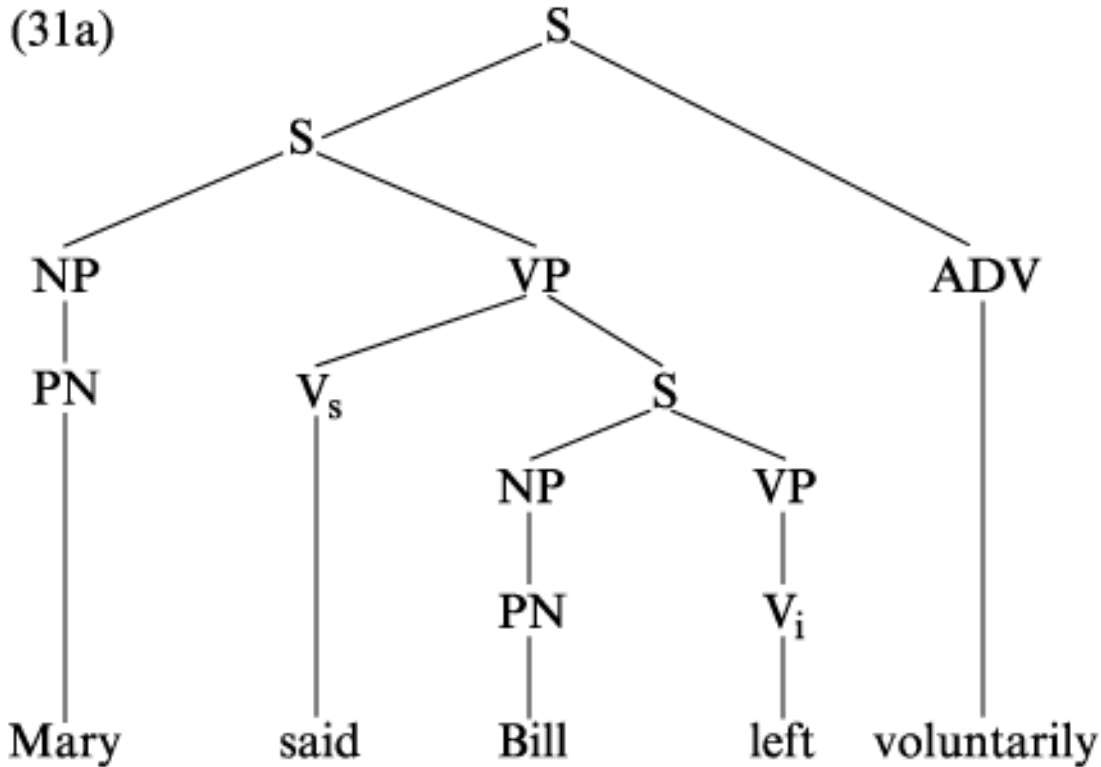
We can implement this suggestion by providing distinct phrase markers corresponding to (4a) and (4b). We will consider two rival ways of doing this. The word ‘voluntarily’ belongs to the category of adverb, so let us add to our lexicon the following:

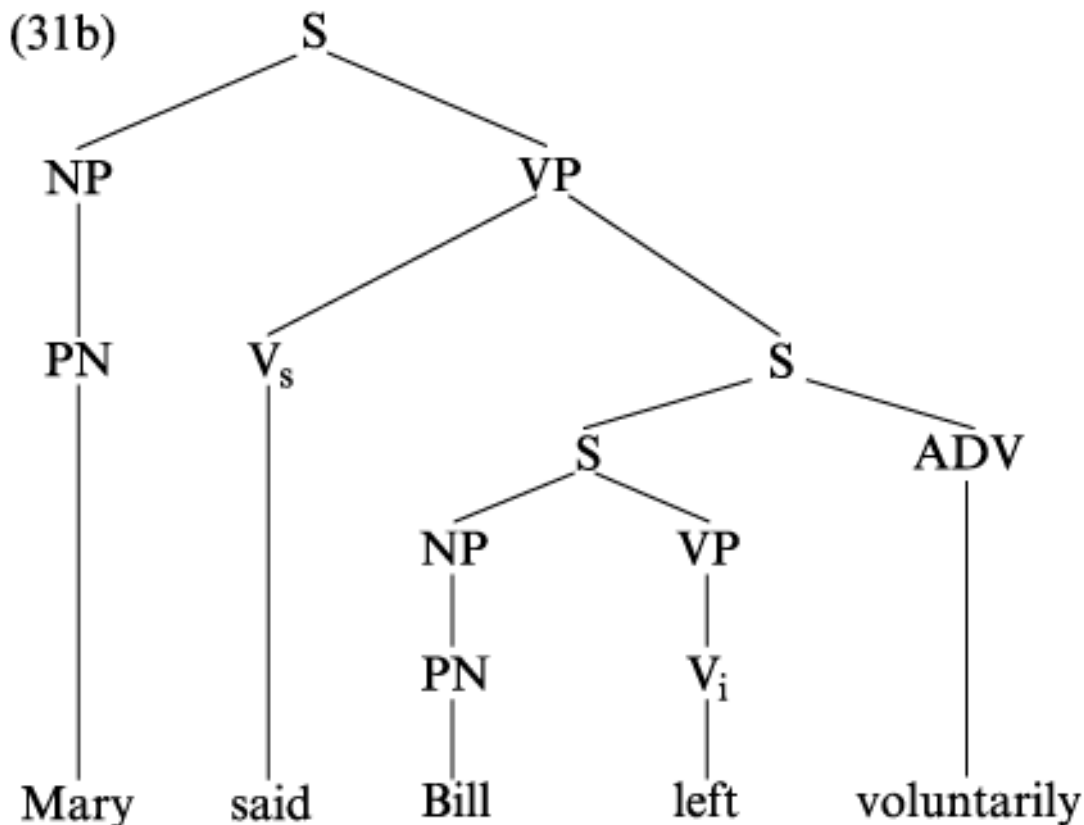
(29) ADV = {voluntarily, hastily, discreetly}.

In line with the informal account of the ambiguity suggested earlier, the first attempt to characterize the syntax of adverbs involves positing the following phrase structure rule:

(30) $S \Rightarrow S + ADV$.

The resulting grammar generates the following distinct phrase markers for (3):





In (31a) the scope - as defined above - of 'voluntarily' is the entire sentence. Thus the verb 'said' is within its scope. By contrast, in (31b) the scope of 'voluntarily' is just the smaller sentence 'Bill left voluntarily'. Thus the verb 'said' is not within its scope, but the verb 'left' is. On the assumption that, as the etymology suggests, an adverb modifies a verb in some way, the verb it modifies is the main verb of the sentence to which it is attached. Thus (31a) corresponds to the reading of (3) that might be loosely paraphrased as 'Of her own volition, Mary said that Bill left' and (31b) corresponds to the reading that might be paraphrased as 'Mary said that Bill left as a matter of his own volition'. (Of course, until an account is provided of how syntactic structure contributes to meaning, the pairing of distinct phrase markers with distinct readings of a string is largely intuitive.)

We now make two observations about our definition of scope. First, by the definition provided, every constituent has a scope; but there are only certain types of constituents whose scopes will interest us (usually, connectives, quantifiers, verbs and adverbs). Second, notice that it is not always the case that for two constituents one is within the scope of the other.

It is worth asking whether this is the best way of characterizing the syntax of sentences containing adverbs. An alternative account emerges if we take the etymology of 'adverb' more seriously and view ADVs as attaching to VP nodes rather than S nodes. Only detailed empirical and logical work will tell which of these analyses is better. The best theory may tell us that there are two distinct types of ADV: one sort that attaches to an S node, another that attaches to VP. There is no need for us to settle this difficult matter here. Suffice to say, that if we find that claims about the semantic properties of adverbs are doing some philosophical work, we will have to look carefully at their syntax.

More than a few philosophical muddles have been due to confusions over the scope of operators in natural language sentences. Utilizing the resources of syntactic theory, it becomes possible for us to study scope relations in a more rigorous fashion, and to marshal arguments for and against certain proposals about the logical form of natural language. Indeed, whenever we encounter philosophical claims which rely upon the semantic properties of quantifiers, adverbs and so on, we will want to look carefully at their syntactic properties. This specific point is just an instance of a quite general principle concerning the relationship between the philosophy of language and theoretical syntax: the former cannot properly proceed independently of the latter.

See also: [Analytical philosophy](#); [Chomsky, N.](#); [Language of thought](#); [Logical form](#); [Mind, computational theories](#)

References and further reading

- Ajdukiewicz, K.** (1935) 'Die Syntaktische Konnexität', *Studia Philosophica* 1: 1-27; trans. 'Syntactic Connexion', in S. McCall (ed.) *Polish Logic*, Oxford: Clarendon Press, 1967. (A full statement of the theory of categorial grammar.)
- Chomsky, N.** (1957) *Syntactic Structures*, The Hague: Mouton. (Early statement of the formal limitations of various approaches to syntax, mathematical properties of different types of grammars, and the theory of transformational grammar. Very readable.)
- Chomsky, N.** (1965) *Aspects of the Theory of Syntax*, Cambridge, MA: MIT Press. (Detailed statement of Chomskian methodology, and Chomsky's 'Standard Theory'.)
- Chomsky, N.** (1981) *Lectures on Government and Binding*, Dordrecht: Foris. (Detailed statement of major components of Chomsky's 'principles and parameters' approach to syntax that moves well beyond the transformational approach of his earlier work. Hard going.)
- Chomsky, N.** (1986) *Knowledge of Language*, New York: Praeger. (Very readable account of the philosophical and technical details of the 'principles and parameters' approach.)
- Davidson, D.** (1980) 'The Logical Form of Action Sentences', in *Essays on Actions and Events*, Oxford: Clarendon Press. (A proposal about the logical form of sentences involving action verbs.)
- Lewis, D.K.** (1983) 'General Semantics', in *Philosophical Papers*, vol. 1, Oxford: Oxford University Press. (Accessible exposition of a categorial grammar for a part of English and a corresponding compositional semantics.)
- May, R.** (1985) *Logical Form: Its Structure and Derivation*, Cambridge, MA: MIT Press. (First sustained attempt to connect work in Chomsky (1981) with traditional philosophical work on 'logical form'.)
- Neale, S.** (1994) 'Logical Form and LF', in C. Otero (ed.) *Noam Chomsky: Critical Assessments*, London: Routledge & Kegan Paul. (An attempt to connect Chomsky's syntactic programme to work on meaning, reference and logical form.)
- Russell, B.A.W.** (1905) 'On Denoting', *Mind* 14: 479-93; repr. in *Logic and Knowledge: Essays 1901-1950*, ed. R.C. Marsh, London: Allen & Unwin, 1956, 41-56. (The classic theory of definite descriptions that provided a paradigm for analysis of logical form.)
- Sells, P.** (1985) *Lectures on Contemporary Syntactic Theories*, Stanford, CA: Center for the Study of Language and Information, Stanford University. (Accessible overview and exposition of competing approaches to generative grammar.)
- Wittgenstein, L.** (1922) *Tractatus Logico-Philosophicus*, trans. D. Pears and B. McGuinness, London: Routledge & Kegan Paul, 1961. (An early, philosophically ambitious theory of logical form.)

Tarski's definition of truth

Alfred Tarski's definition of truth is unlike any that philosophers have given in their long struggle to understand the concept of truth. Tarski's definition is more clear and precise than any previous definition, but it is also unusual in character and more restricted in scope. Tarski does not provide a general definition of truth. He provides instead a method of constructing, for a range of formalized languages L , definitions of the notions 'true sentence of L '. A remarkable feature of Tarski's work on truth is his 'Criterion T', which lays down a general condition that any definition of 'true sentence of L ' must satisfy. Tarski's ideas have exercised an enormous influence in philosophy. They have played an important role in the formulation and defence of a range of views in logic, semantics and metaphysics.

1 Formal correctness and material adequacy

Any definition that aims to elucidate some aspect of a concept must satisfy two sets of requirements: the definition must be, in Tarski's terminology, 'formally correct' and 'materially adequate'. Under formal correctness fall requirements concerning the form of the definition and the vocabulary that the definition may use. Concerning these requirements, let us note only that a definition of 'true' of the form

x is true iff x

is bound to be formally correct provided that the right-hand side of the equivalence contains only 'admissible' vocabulary, that is, vocabulary that may legitimately be used in defining 'true'. For example, the definition

x is true iff x corresponds to reality

fails to satisfy the requirement of formal correctness, because the terms 'corresponds' and 'reality' are much too unclear to be considered admissible. On the other hand, no doubts should arise about the formal correctness of the definition

x is true iff x is identical to x .

The definition is, of course, unsatisfactory. But that is because it is materially inadequate: it fails to conform to our ordinary notion of truth.

The most remarkable feature of Tarski's work on truth (1936, 1944) is his formulation, known as 'Criterion T', of the requirement of material adequacy. It is this Criterion that is responsible for the special characteristics, both positive and negative, of Tarski's definition. On the positive side, it enables Tarski to give a precise definition of truth and to prove its material adequacy. On the negative side, it is the source of the unusual and, from the philosophical point of view, strange character of Tarski's definition.

Criterion T lays down two conditions that a materially adequate definition of truth must meet. The second, and less important, condition is that the definition should imply that only sentences are true. Tarski imposes this condition because he takes truth to be a predicate of sentences. The first, and more important, condition is that the definition should imply all the 'T-biconditionals', that is, all sentences of the form

(T) ' ϕ ' is a true sentence (of English) iff ϕ .

Thus, Criterion T requires that a definition of truth (for English) should imply, for example, the famous T-biconditional,

'Snow is white' is a true sentence (of English) iff snow is white.

Note that any definition that implies this biconditional is bound to define a concept that agrees with truth over the sentence 'Snow is white': the sentence will fall under the concept if and only if it falls under truth. Hence, any definition that implies *all* the T-biconditionals is bound to define a concept C that agrees with truth over all the sentences: the concept defined is bound to be coextensive with truth (assuming that truth and C agree over non-sentences). Furthermore, since this argument uses no special contingent assumptions, the claim of

coextensiveness holds irrespective of the contingent facts. In one sense of 'intension', then, the concept defined is bound to have the same intension as truth.

The equivalence of the two sides of the T-biconditionals has been noted by many philosophers and logicians, including Aristotle, Gottlob Frege and Frank Ramsey. Tarski is the first, however, to have used the equivalence to formulate a material adequacy condition on a definition of truth. The simplicity of Tarski's formulation should not obscure the important and surprising claim that it contains. Definitions of virtually all of our concepts can be assessed for material adequacy only in a piecemeal way, namely, by comparing their consequences with the ordinary uses of the concepts defined. No general conditions of material adequacy can be laid down. One would have expected the same of truth. But Criterion T puts forward a general and a priori condition of material adequacy on a definition of truth - a condition that, surprisingly, is highly plausible. Criterion T embodies Tarski's key philosophical claim. If it is granted then there can be no denying that it marks an important philosophical advance.

The formulation of Criterion T given above, it must be acknowledged, is unsatisfactory in several respects. First, it is formulated only for a definition of truth *for English in English*. What is wanted - and what Tarski's works provide - is a more general criterion that also applies to other languages. Second, the criterion as formulated is plausible only for sentences that are free of ambiguity, indexicals, tenses and other context-dependent elements. Tarski himself never attempted to overcome this limitation - his primary interest was in formalized languages for which these problems do not arise - but other philosophers have attempted to do so. Third, and most important, the criterion as formulated requires the definition of truth to be inconsistent. This inconsistency is a consequence of the liar paradox.

2 The liar paradox

Suppose, following Saul Kripke, that we introduce a name 'Jack' into English with the understanding that it names the sentence 'Jack is not a true sentence (of English)'. Suppressing the parenthetical relativization, the T-biconditional for Jack is

'Jack is not a true sentence' is a true sentence iff Jack is not a true sentence.

Our stipulation ensures that

Jack = 'Jack is not a true sentence'.

Hence, by the principle of substitutivity of identicals, the T-biconditional for Jack implies

Jack is a true sentence iff Jack is not a true sentence.

This in turn yields a contradiction by the rules of propositional logic. Note that the difficulty cannot be overcome by stipulating away names such as 'Jack'. Self-referential 'liar' sentences, which yield an inconsistency, can be constructed without the aid of such names.

Tarski makes constructive use of the argument of the liar paradox. He uses it to prove (a version of) the indefinability theorem named after him: arithmetical truth is not arithmetical. More precisely, no formula of Peano arithmetic (PA) is true precisely of (the codes of) the true sentences of PA (see **Gödel's theorems**). The theorem has a more general version: no classical language L_1 that is rich in syntactic resources can contain its own notion of truth. L_1 can be extended to a new language L_2 by the addition of a predicate that represents 'true in L_1 '. But, by Tarski's theorem, L_2 will not contain its own truth predicate. As before, L_2 can be extended to yet another language L_3 ; this will contain a truth predicate for L_2 , but again not for L_3 . By repeating this procedure, we obtain a Tarskian hierarchy of languages, $L_1, L_2, \dots, L_n, \dots$, in which truth for a language of level n is contained in the language at level $n + 1$. 'Tarski's indefinability theorem' motivates a way of getting around the difficulty that the paradoxes pose for Criterion T: (1) Draw a distinction between the language *in* which the definition of truth is given (the 'metalanguage') and the language *for* which the definition is given (the 'object language'); and (2) require that the object language not contain its own notion of truth. This manoeuvre ensures that no liar sentences occur in the object language. Consequently, the threat of inconsistency is removed from the T-biconditionals, and we regain Criterion T as a plausible material adequacy condition.

The manoeuvre allows us to proceed with the definition of truth for certain kinds of languages. But it does not help

us to understand the behaviour of truth in languages (such as English) that contain paradoxical sentences.

Tarski's response to this problem was an impatient one: he dismissed all such languages as inconsistent. In the past twenty-five years or so, however, Tarskian ideas have been used to develop a better, and certainly less impatient, response to the problem. The response that is most Tarskian in spirit is the 'hierarchy theory' of Charles Parsons and Tyler Burge. This theory sees our ordinary notion of truth as implicitly hierarchical. Each occurrence of 'true', according to the theory, corresponds to a truth predicate at a certain level in Tarski's hierarchy - with different occurrences corresponding to predicates at possibly different levels. So, for example, in the sentence

“‘Snow is white’ is true’ is true.

the first occurrence of 'true' might correspond to the truth predicate at the first level of the hierarchy, and the second occurrence to that at the second level. The substantial burden on a hierarchy theory is to give a systematic explanation of the levels of the occurrences of 'true'. There are other approaches to the paradoxes that build on Tarski's ideas. The principal ones are the 'fixed-point theory' (see [Semantic paradoxes and theories of truth §4](#)) and the 'revision theory' (see [Definition §4](#)).

Tarski showed how to construct definitions of truth that meet Criterion T. The next section contains some examples of Tarskian truth definitions. But before turning to these, let us note some general features of Tarski's method. First, this method does not yield a definition of the relational notion 'true in L ', with variable L ; its primary application is to the definition of 'true in L ' for particular L . Second, the method enables us to give truth definitions only for formalized languages, not for natural languages. Third, it yields a definition in a metalanguage ML that is distinct from the object language L . If ML is a classical language then the distinction is forced by Tarski's indefinability theorem.

3 Examples of Tarskian truth definitions

This section gives three examples of Tarskian truth definitions. Two of these are very simple; only the third begins to approach the complexity of Tarski's actual definition.

Example A. Suppose that L is a fragment of English that contains only a finite number of sentences (say, a million). Suppose that the sentences of L contain no occurrence of 'true' or of any other term whose use is inadmissible in a definition of truth. Further, let ' S_1 ', ' S_2 ', ..., ' S_{10^6} ' be abbreviations of the sentences of L . A definition of 'true sentence of L ' may now be given as follows:

x is a true sentence of L iff S_1 and x is ' S_1 ', or S_2 and x is ' S_2 ', or ..., or S_{10^6} and x is ' S_{10^6} '.

The definition is evidently formally correct. Apart from the vocabulary of $S_1, S_2, \dots, S_{10^6}$, which is stipulated to be admissible, it uses only the logical notions 'or' and 'is' (identity) and the quotation names for the sentences of L . The definition is also materially adequate since, under the assumption of such simple syntactic facts as ' S_1 ' is not ' S_2 ', it implies all instances of the form (T).

Example B. Let us now construct a definition of truth for a language L' that has an infinite number of sentences. Suppose that L' has two sentential connectives, ' $\sim$ ' (negation: 'it is not the case that') and 'Čamp;' (conjunction: 'and'). Suppose further that L' has a finite number of atomic sentences, that is, sentences that are not built using the connectives ' $\sim$ ' and 'Čamp;'. Now truth for atomic sentences can be defined as in example A. Over compound sentences, truth can be defined recursively: the negation of a sentence A is true iff A is not true; a conjunction is true iff both its conjuncts are true. Putting these elements together we can construct an explicit definition of 'true sentence of L' ' in the following way. First let us define an intermediary notion, that of a 'closed' set; this will enable us to give a more compact definition of truth. A set Z is 'closed' iff (1) Z contains all true, and only true, atomic sentences of L' ; (2) Z contains the negation of a sentence A iff it does not contain A ; and (3) Z contains the conjunction of sentences A and B iff it contains the conjuncts A and B .

x is a true sentence of L' iff x belongs to every closed set.

Assuming that the atomic sentences of L' do not contain any terms that are inadmissible, our definition is formally correct. Further, it is materially adequate, since it implies all the T-biconditionals of L' . Note that the implication requires some syntactic facts about L' and some low-level set-theoretic facts.

Example C. Finally, let us see how to construct a definition of truth for a quantificational language L'' . Such a language has an infinite number of sentences, so the definition of truth cannot proceed in the list-like way of example A; it relies instead on a recursion. However, in these languages compound sentences may have immediate components that are not themselves sentences. For example, the quantificational compound ' $(\forall v)Red(v)$ ' ('Everything is red') has as its immediate component ' $Red(v)$ ', which is not a sentence. Consequently, in these languages, the truth of compounds cannot in general be explained recursively in terms of the truth of their components. To solve the problem Tarski first defines recursively a more general notion, namely, 'satisfaction', and then he defines truth using satisfaction.

Satisfaction is a relation that holds between 'assignments' (of values to variables) and formulas. We can think of assignments as functions from variables into objects. So, if s is an assignment and v is a variable then $s(v)$ is the value s assigns to v . Intuitively, an assignment s satisfies a formula A iff A is true when the free variables in it are interpreted as naming the values that s assigns them. The notion of satisfaction is easy to define for atomic formulas: any assignment that assigns a red object to the variable v satisfies the formula ' $Red(v)$ '. Similarly, any assignment s such that $s(v)$ is larger than $s(w)$ satisfies the formula ' $Larger(v, w)$ '. The clauses for n -ary predications are analogous. Satisfaction for complex formulas is defined recursively in terms of the satisfaction of the components:

An assignment s satisfies the negation of a formula A iff s does not satisfy A .

An assignment s satisfies the conjunction of B and C iff s satisfies B and s satisfies C .

An assignment s satisfies the universal quantification of A with respect to the variable v iff A is satisfied by all assignments s' such that for all variables v' distinct from v , $s'(v') = s(v')$.

This recursive explanation can be converted to an explicit definition of satisfaction (see example B). It is easy to show that the satisfaction of a formula A by an assignment s depends only on the values that s assigns to the free variables of A . Consequently, a sentence is satisfied either by all assignments or by none. This motivates the following definition of truth:

x is a true sentence of L'' iff x is a sentence of L'' that is satisfied by all assignments.

The reader will find a more leisurely presentation of this example in W.V. Quine's *Philosophy of Logic* (1970).

4 Philosophical influence

Tarski's avowed aim in constructing a definition of truth was to lay down the foundations of a scientific semantics, and it is here that his work has been most influential. Tarski's work provides the foundations for a semantic study of logic - a study that has been vigorously pursued in the twentieth century and to which Tarski made important contributions. Two key notions in this study - 'model' and 'truth in a model' - are generalizations of the notions of 'assignment' and 'satisfaction' introduced in example C. An assignment specifies the values of the variables; a model specifies the domain of quantification and the 'values' of all the nonlogical constants in the language. Furthermore, just as satisfaction tells us what formulas are true under what assignments, similarly 'truth in a model' tells us what formulas are true when the quantifiers and the nonlogical constants are interpreted as specified in the model. The notions 'model' and 'truth in a model' enable a semantic explanation of the various logical notions. For example, an argument is 'valid' if and only if its conclusion is true in all models in which all the premises are true. The semantic analysis of logical concepts has yielded valuable insights into logic (see [Model theory](#)).

Tarski's definition of truth is also of importance in natural-language semantics. Here it has inspired two different kinds of approaches: the Davidsonian approach and the model-theoretic approach. The Davidsonian approach takes the goal of a semantic theory for a language L to be the construction of a theory of truth for it: a finite theory that implies all the T-biconditionals of L . (This is only a rough characterization; see [Meaning and truth](#).) Davidson has argued that such a theory must proceed recursively, explaining semantic properties of compounds in terms of the semantic properties of the components. As a result, the theory will reveal the logical structure of language. The model-theoretic approach, developed by Richard Montague and others, takes the goal of the semantic theory to be the definition of the notion 'truth in a model'. This requires the characterization of the notion of model and it, in turn, requires an analysis of the logical structure of language. Davidson's work has been highly influential in

philosophy, but less so in linguistics. The model-theoretic approach has, however, gained many adherents in linguistics.

Tarski gave his definition in a philosophical climate that was sceptical of the notion of truth. Some logical positivists suggested, for example, that the notion was unanalysable in empirical terms and, hence, unacceptable in a scientific philosophy. Tarski's definition provided a decisive refutation of their suggestion. Let L be the positivists' language - purified of all elements that the positivists judge to be unempirical and meaningless. The notion 'true in L ', Tarski showed, can be rigorously defined using only resources that the positivists would accept: the definition uses only some syntactic and logical concepts and those that are expressed in L . Tarski's results were accepted by the positivists and significantly influenced their philosophical work (especially that of Rudolf Carnap).

Tarski's definition has also been influential in more recent philosophy. Some philosophers (for example, Hartry Field (1972) and Michael Devitt) have sought to construct a correspondence theory of truth on its basis. These philosophers accept the recursive parts of Tarski's definition. But they argue that the definition does not provide a substantial account of the semantic properties of the simplest elements of the language. It does not explain, for example, when a name denotes an object or when an object falls under a predicate. To obtain a fully fledged correspondence theory, they suggest, Tarski's recursive explanation of truth should be supplemented with a substantial account of the semantics of the simples. Some philosophers have seen in this kind of theory a way of vindicating realism. But others have questioned both the feasibility of the theory and its links to the realism issue.

It is ironic that while Tarski put forward his definition to show that the concept of truth is useful for certain logical and philosophical purposes, his ideas have also been used in defence of 'deflationism' - the view that the concept has no substantive role to play in logic and philosophy. The thought here is that something like Tarski's definition explains completely the meaning of 'true', and that this makes truth a philosophically lightweight concept. Hence, it is claimed, the concept can play no substantive role in, for example, the explanation of validity or the realism debate. Consider, for instance, the explanation of validity - say, of the argument P , therefore Q - in terms of truth: the argument is valid iff it is impossible that P be true but Q false. The T-biconditionals, the deflationists argue, reduce this explanation to the nearly vacuous claim that the argument is valid iff it is impossible that P and not Q . Neither the explanation nor the concept of truth seem to be doing any substantive work.

The role of the concept of truth in philosophy continues to be a topic of intense debate. In this debate, Tarski's definition and Tarski's ideas occupy centre stage.

See also: [Consequence, conceptions of §5](#); [Logical and mathematical terms, glossary of](#); [Semantics](#)

ANIL GUPTA

References and further reading

- Etchemendy, J.** (1988) 'Tarski on Truth and Logical Consequence', *Journal of Symbolic Logic* 53: 51- 79. (Nontechnical; makes useful critical points.)
- Field, H.** (1972) 'Tarski's Theory of Truth', *Journal of Philosophy* 69: 347-75; repr. in M. Platts (ed.) *Reference, Truth and Reality*, London: Routledge & Kegan Paul, 1980.(Referred to in §4; argues for a transformation of Tarski's definition into a correspondence theory of truth.)
- Gupta, A. and Belnap, N.D.** (1993) *The Revision Theory of Truth*, Cambridge, MA: MIT Press. (Chapters 1 and 4 are nontechnical and discuss Criterion T and the liar paradox; the rest of the book is quite technical.)
- Horwich, P.** (1990) *Truth*, Oxford: Blackwell.(A clear defence of deflationism.)
- Quine, W.V.** (1970) *Philosophy of Logic*, Englewood Cliffs, NJ: Prentice-Hall.(Referred to in §3; presents a Tarskian truth definition in detail.)
- Tarski, A.** (1936) 'Der Wahrheitsbegriff in den formalisierten Sprachen', *Studia Philosophica* 1: 261-405; trans. J.H. Woodger, 'The Concept of Truth in Formalized Languages', in *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*, Oxford: Clarendon Press, 1956; ed. J. Corcoran, Indianapolis, IN: Hackett Publishing Company, 2nd edn, 1983, 152-278.(The classic paper on the subject; the introduction and the first section of the paper are nontechnical and are highly recommended.)
- Tarski, A.** (1944) 'The Semantic Conception of Truth and the Foundations of Semantics', *Philosophy and Phenomenological Research* 4: 341-76; repr. in L. Linsky (ed.) *Semantics and the Philosophy of Language*, Urbana, IL: University of Illinois Press, 1952.(Nontechnical; includes valuable philosophical material.)

Tarski, A. (1969) 'Truth and Proof', *Scientific American* 220 (6): 63-77; repr. in R.I.G. Hughes (ed.) *A Philosophical Companion to First-Order Logic*, Indianapolis, IN: Hackett Publishing Company, 1993. (A particularly good popular exposition of the main ideas.)

Tense and temporal logic

A special kind of logic is needed to represent the valid kinds of arguments involving tensed sentences. The first significant presentation of a tense logic appeared in Prior (1957). Sentential tense logic, in its simplest form, adds to classical sentential logic two tense operators, *P* and *F*. The basic idea is to analyse past and future tenses in terms of prefixes 'It was true that' and 'It will be true that', attached to present-tensed sentences. (Present-tensed sentences do not need present tense operators, since 'It is true that Jane is walking' is equivalent to 'Jane is walking'.) Translating the symbols into English is merely a preliminary to a semantics for tense logic; we may translate '*P*' as 'it was true that' but we still have the question of the meaning of 'it was true that'. There are at least two versions of the tensed theory of time - the minimalist version and the maximalist version - that can be used for the interpretation of the tense logic symbols.

The minimalist version implies that there are no past or future particulars, and thus no things or events that have properties of pastness or futurity. What exists are the things, with their properties and relations, that can be mentioned in certain present-tensed sentences. If 'Jane is walking' is true, then there is a thing, Jane, which possesses the property of walking. 'Socrates was discoursing', even if true, does not contain a name that refers to a past thing, Socrates, since there are no past things. The ontological commitments of past and future tensed sentences are merely to propositions, which are sentence-like abstract objects that are the meanings or senses of sentences. 'Socrates was discoursing' merely commits us to the proposition expressed by the sentence 'It was true that Socrates is discoursing'.

The maximalist tensed theory of time, by contrast, implies that there are past, present and future things and events; that past items possess the property of pastness, present items possess the property of presentness, and future items possess the property of being future. 'Socrates was discoursing' involves a reference to a past thing, Socrates, and implies that the event of Socrates discoursing has the property of being past.

1 *De dicto* and *de re* occurrences of tenses

If a tense occurs *de dicto*, then the tensed expression ascribes a temporal property to a 'dictum' (proposition). 'It will be true that the sun is exploding' involves a *de dicto* occurrence of 'will be': the property 'having future truth' is ascribed to the proposition expressed by 'The sun is exploding'. If a tense occurs *de re*, then the tensed expression ascribes a property to a thing or event. The sentence 'The sun will explode' involves a *de re* use of 'will' since the tensed expression 'will explode' ascribes the temporal property 'exploding in the future' to the sun (see *De re/de dicto*).

The tense logic favoured by minimalists is based on the idea that all occurrences of past and future tenses are *de dicto* (even though some might appear to be *de re*) and therefore can be represented by sentential operators, such as 'it was true that'. According to this theory, the sentence 'The sun will explode' only appears to ascribe a temporal property to the sun; in reality it ascribes a temporal property to a proposition. 'The sun will explode' means the same as 'It will be true that the sun is exploding'.

2 Quantified tense logic

The *de re/de dicto* distinction also appears in the different approaches of the maximalist and minimalist theorists to the syntax of quantified tense logic. Quantified tense logic may be formulated in at least two ways. One way is based on the idea that quantifiers ('any', 'some') occur only within the scope of tense operators. In the sentence 'Some thing will fly to Mars', the quantifier 'some' does not occur within the scope of a tense operator. But it does in 'It will be true that some thing flies to Mars'.

The minimalist theory implies that quantification over past or future particulars should occur only in the scope of tense operators. Consider the sentence 'Some person did exist who wrote *The Republic*'. The minimalist will take this to express the same proposition as 'It was true that some person is writing *The Republic*', which has the symbolic form $P(\exists x)Gx$. The quantified expression 'some person' occurs within the scope of 'it was true that'; this implies that 'some person' does not now refer to anything (there are no past people for it to refer to), but used to refer to something; to a person that existed at the time the proposition was true.

The minimalist assumption that temporal properties should be ascribed only to propositions also appears in the treatment of quantification over present particulars. The sentence ‘Some person is presently writing a book’ is not interpreted as ascribing the temporal property ‘presently writing a book’ to a person. Rather, it is taken to mean ‘It is true that some person is writing a book’, which would involve a present tense operator. This sentence ascribes the property of being presently true to the proposition ‘Some person is writing a book’. This minimalist interpretation of quantification is exemplified in A.N. Prior’s remark in *Past, Present and Future*: ‘a quantifier preceding any such operator [any tense operator] is naturally taken to be governed by the “It is the case that-”, which is prefixable to anything we say, and therefore to range over what *now* exists’ (1967: 144; *emphasis added*).

The maximalist tensed theory of time, by contrast, allows quantification over past, present or future particulars to occur outside the scope of tense operators. The expression ‘some person’ in the sentence ‘Some person existed who wrote *The Republic*’ quantifies over all past, present and future persons. This reflects the assumption that some people exemplify pastness, some, presentness and some, futurity.

The maximalist theory implies a *de re* use of the symbols P and F . Traditional tense logic (following Prior) uses P and F only in a *de dicto* manner, to mean ‘it was true that’ or ‘it will be true that’. But in an expression of the form $(\exists x)Px$, ‘ P ’ does not mean ‘it was true that’ but stands for a property of x ; the property of pastness. The expression translates as ‘Something is past’. Since ‘ P ’ here is not used as a sentential operator, but as a predicate, this should be marked in our symbolic notation. We may put an asterisk after P to mark the use of this symbol as a predicate. Thus we should say $(\exists x)P^*x$.

The symbol P is also used in a *de re* manner in $(\exists x)[PG]x$. Here the symbol P operates on the predicate G to form a more complex predicate. This translates as ‘Something exemplifies past G -ness’ or, equivalently, ‘Something’s exemplification of G has pastness’. To mark this distinctive use of P as a predicate operator, we can underline P , so that we say $(\exists x)[\underline{P}G]x$.

According to the maximalist tensed theory of time, an adequate quantified tense logic must include P (a sentential operator), P^* (a predicate) and $\underline{P}$ (a predicate operator). Given these symbolic distinctions, maximalists can allow that P is used only *de dicto*.

The distinction between *de dicto* and *de re* occurrences of tense logic symbols is partly analogous to the same distinction in modal logic. The box is used *de dicto* in $\Box A$, which means ‘It is necessarily true that A ’, where A is some proposition. The box is used *de re* in $(\exists x)\Box Fx$, which means that some thing is necessarily F .

3 A paradox for the minimalist theory

The minimalist tensed theory of time is false if there is a true sentence which a minimalist interpretation turns into a logical contradiction. The maximalist may claim there are such sentences. For example, the maximalist may claim that 10 billion years ago, before any intelligent organisms evolved, there was no language and thus no truths or falsehoods. Consider the sentence ‘It is logically possible that there was a time t at which there were no truths’. If ‘there was’ and ‘there were’ are *de dicto* occurrences of tense, then this sentence says the same as ‘It is logically possible that it was true at time t that there are no truths’, which is a logical contradiction. The sentence must instead be taken to mean that it is logically possible that some time t exemplifies both pastness and having contained no truths.

The minimalist may counter that Platonic realism is true and thus that it is necessary that there be truths at all times. The maximalist may respond that the truth of Platonic realism and the falsity of nominalism are determined by metaphysics, not by formal logic. ‘Platonic realism is true’ is not a logical theorem and ‘There were no truths 10 billion years ago’ is not a logical contradiction. Thus a formal logic that implies these sentences are logical truths or falsehoods is not in fact valid.

The minimalist may have a second response, namely that the sentence ‘It is logically possible that it was true at time t that there are no truths’ does not imply the existence of a truth at time t . The tense operator ‘it was true at time t that’ does not imply that the proposition ‘There are no truths’ *existed* at time t . Rather, it implies merely that the time at which the (present tense) proposition should be evaluated for truth or falsity is the past time t , and that the proposition possesses truth at this time. The maximalist will say that this response depends on the intelligibility of the thesis that the proposition ‘There are no truths’ possesses at time t the property of being true and yet does

not exist at this time. How can a nonexistent (at time t) proposition possess a property (at time t)? It is debatable whether the maximalist can derive a contradiction from the minimalist theory that some propositions do not exist at some of the times at which they exemplify the property of being true.

A.N. Prior, who presented the first significant tense logic (1957), defends a minimalist tensed theory of time (1967) and is the most influential opponent of the maximalist theory and *de re* occurrences of tenses. Prior's critical stance is analogous to Quine's position regarding quantified modal logic. Prior found problematic *de re* occurrences of 'was' and 'will be', objecting to the realism about properties of pastness, futurity and presentness associated with constructions of the form $(\exists x)P^*x$ (something exemplifies 'being past'). (Similarly, Quine found problematic *de re* occurrences of 'must be', and objected to the realism about essential properties associated with constructions of the form $(\exists x)\Box Fx$ (something exemplifies 'being necessarily F ').)

In Prior's case, the rejection of the realist position is in part due to his categorization of pastness, presentness and futurity as *activities*. Prior suggests that these temporal properties of events (if there were such properties) belong in the same ontological category as activities done by things. He denies that there are *events* 'momentarily doing something called "being present" and then doing something else called "being past" for much longer' (1967: 18). Prior may be interpreted as suggesting that it is a mistake to categorize presentness and pastness along with such activities as running or talking, and then say that presentness and pastness (if there are such properties) are activities performed not by things, but by events.

The realist may grant this point, but argue that presentness and pastness need not be categorized as activities performed by events. 'Events possess temporal properties of being present or being past or being future' does not entail 'Presentness, pastness and futurity are activities done by events'. These temporal properties, like such properties as existence or self-identity, are unusual kinds of properties and should be placed in an ontological category by themselves.

4 Do *de dicto* occurrences imply *de re* occurrences?

The idea that all occurrences of tenses are *de dicto* leads to a minimalist semantic interpretation of such present-tensed sentences as 'The moon is bright'. According to this interpretation, this sentence does not report that an event, the moon's exemplification of brightness, has presentness. There is no first-order property of presentness and no event; there is only the thing, the moon, exemplifying brightness. The maximalist may argue that problems with this interpretation arise as soon as we ask *when* the moon exemplifies brightness. If 'exemplify' is tenseless, then there is no answer, and the analysis is faulty, since the present tense sentence 'The moon is bright' implies the moon is now exemplifying brightness. But if 'exemplify' is present-tensed, then we must ask what the semantic content of the present-tensed aspect of this verb is. The present tense must have some semantic content or meaning, since it conveys temporal information and is not a meaningless syntactic device such as 'It' in 'It is raining'. Further, to what ontological category does this temporal information belong?

According to the minimalist *de dicto* theory of tenses, the semantic category of the present tense is 'sentential operator' (the present tense has the same semantic content as 'it is true that') and the relevant ontological category is 'second-order property'; the present tense ascribes the property of 'being presently true' to a proposition. This theory implies that if the sentence 'The moon is bright' is true, we are ontologically committed to (1) the proposition, 'The moon is bright', (2) the proposition's exemplification of the property, 'being presently true', (3) the thing, the moon, and (4) this thing's exemplifying the property of brightness. The maximalist may object that 'exemplifying' in (4) is present-tensed, reflecting the fact that the thing presently exemplifies a property of brightness. According to the maximalist, the thing's *present* exemplification of brightness (rather than a timeless, past or future exemplification of brightness) is the ground for the proposition's being presently true.

5 Past and future existents

According to the minimalist theory, the ontological commitments of the sentence 'Plato had been alive' are merely the proposition 'Plato is alive' and the proposition's property of having been true. The maximalist may grant that these are the only ontological commitments if we accept the coherentist or eliminativist theory of truth. But the maximalist accepts the correspondence theory of truth and argues that this theory requires further ontological commitments. 'Plato had been alive' commits us to a past thing, namely, Plato, and a past event or state, namely,

Plato's being alive. 'Correspondence' is a symmetrical property. (It is a dyadic property and thus a relation.) If the proposition 'Plato is alive' had been true in the past, then this proposition exemplifies past correspondence to Plato's being alive. Since correspondence is a symmetrical relation, this implies that some event, Plato's being alive, exemplifies past correspondence to the proposition 'Plato is alive'. The reason this event exemplifies past correspondence is that the event exemplifies pastness. Since Plato is a part of this past event, Plato exemplifies pastness. It follows, the maximalist concludes, that some events and things are past.

This maximalist conclusion implies that the temporal property 'pastness' is *de re*, since it is a property of a thing, Plato, and of an event, Plato's being alive. There was something, Plato, over which we can now quantify; it is the case that $(\exists x)P^*x$. The same results hold for propositions that will correspond to something exemplifying some property.

One concern that minimalists have about the maximalist interpretation of tense logic pertains to the maximalists' thesis that past or future things or events *presently* possess properties. Consider the formula $(\exists x)P^*x$. Does not $(\exists x)$ in this formula convey that there presently exists an x ? And does not P^*x mean that this x is not presently existing but is past? If so (the minimalist may argue), the symbolic representation of *de re* occurrences of tenses turns true sentences into contradictions.

But the maximalist may respond that $(\exists x)$ has at least three different meanings in an adequate quantified tense logic, corresponding to three different senses of 'exists'. If 'exists' is present-tensed, it can have one of two senses. In one sense, it is logically equivalent to 'has the property of presentness', so that 'Mount Everest exists' is logically equivalent to 'Mount Everest has the property of presentness'. In the second sense, it can be equivalent to 'presently possesses some property'. Something that is past or future exists in this present-tensed sense, since if something is now past, it presently possesses the property of pastness. If 'exists' is tenseless, it is in effect disjunctively tensed; that is, it is equivalent to 'has existed, exists or will exist', where the middle 'exists' is equivalent to 'has the property of presentness'. If we say 'Socrates exists', we mean that Socrates either has existed or exists (presently) or will exist.

The maximalist argues that this threefold distinction shows that there is no logical difficulty in the idea that some past or future item presently possesses a property. The maximalist may note that 'Some past thing presently has the polyadic property of being remembered' is not a logical contradiction. Nor can it be turned into a logical contradiction by substitution of synonyms. Indeed, the maximalist continues, an analysis of the concept expressed by 'remembering' implies that the minimalist theory is false. Remembering is not a relation between a present person and a present image. If it were such a relation, it would follow that what is being remembered is in every case an image that exists simultaneously with the act of remembering. But this is an analytic contradiction, since it is a part of the logic of 'remembers' that what is remembered exists earlier than the act of remembering. Remembering a thing is a three-termed relation: a person remembers a past thing by forming a present image of that past thing. The 'of' expresses the relation in which the present image stands to the past thing.

The maximalist may elaborate that there will appear to be a difficulty about non-present items presently possessing properties if one is mistaken about the sorts of properties a past or future item can possess. Since Plato does not now exist, Plato cannot now possess the properties of walking, breathing, writing, thinking and so on. These properties can only be possessed by Plato when he is present. But Plato now possesses the properties of being past, of being referred to by the name 'Plato', of being thought about, of being earlier in time than Descartes, and so on. Further, Plato now possesses the properties of having walked, having written *The Republic*, having been alive, having been human, and so on, all of which can be expressed in a symbolic notation of the form $(\exists x)[\underline{P}G]x$, where $(\exists x)$ can be taken as tenseless or as present-tensed in the second sense distinguished by the maximalist.

See also: [Adverbs](#); [Continuants §2](#); [Demonstratives and indexicals §1](#); [Intensional logics](#); [Logical and mathematical terms, glossary of](#); [Modal logic](#); [Modal operators](#); [Time](#)

QUENTIN SMITH

References and further reading

- Burgess, J.** (1984) 'Basic Tense Logic', in D. Gabbay and F. Guenther, *Handbook of Philosophical Logic*, vol. 2, Dordrecht: Reidel. (A technical account of sentential tense logic.)
- Cocchiarella, N.** (1984) 'Philosophical Perspectives on Quantification in Tense and Modal Logic', in D. Gabbay

- and F. Guenther, *Handbook of Philosophical Logic*, vol. 2, Dordrecht: Reidel. (A presentation, with technical detail, of issues in quantified tense logic.)
- Craig, W.L.** (1996a) 'Tense and the New B-Theory of Language', *Philosophy* 71: 5-26. (Presents a defence of the tensed theory of time against the new tenseless theory (the B-theory) of time.)
- Craig, W.L.** (1996b) 'The New B-Theory's *tu quoque* Argument', *Synthese* 105: 1-21. (Shows how recent work on indexicals, such as Kaplan's and Perry's, supports the tensed theory of time.)
- Forbes, G.** (1984) *The Metaphysics of Modality*, Oxford: Clarendon Press. (Includes a clear discussion of tense logic and its relation to modal logic.)
- Gale, R.** (1968) *The Language of Time*, New York: Humanities Press. (An influential book that includes a classic defence of the maximalist tensed theory of time.)
- Haack, S.** (1978) *Philosophy of Logics*, Cambridge: Cambridge University Press. (Chapter 9 of this book includes one of the most accessible introductions to tense logic.)
- Kamp, J.A.W.** (1971) 'Formal Properties of "Now"', *Theoria* 37: 227-73. (An influential technical account.)
- Le Poidevin, R.** (1991) *Change, Cause and Contradiction: A Defence of the Tenseless Theory of Time*, New York: St Martin's Press. (Includes an accessible defence of the thesis that the tensed theory of time is false.)
- Le Poidevin, R.** (ed.) (1998) *Question of Time and Tense*, Oxford: Clarendon Press. (A collection of essays by Le Poidevin, Craig, Nerlich, Oaklander, Dyke, Smith and others about the tensed theory of time.)
- Oaklander, L.N.** (1984) *Temporal Relations and Temporal Becoming*, Lanham, MD: University of America Press. (Presents a sustained criticism of the tensed theory of time.)
- Oaklander, L.N. and Smith, Q.** (eds) (1994) *The New Theory of Time*, New Haven, CT: Yale University Press. (Essays by Kaplan, Oaklander, Mellor, Schlesinger, C. Williams, Smith and others on the tensed theory of time versus the new tenseless theory of time.)
- Prior, A.N.** (1957) *Time and Modality*, Oxford: Clarendon Press. (This classic is the primary source for modern tense logic.)
- Prior, A.N.** (1967) *Past, Present and Future*, Oxford: Clarendon Press. (A collection of papers on time and tense logic, some of which articulate and defend the minimalist tensed theory of time.)
- Prior, A.N.** (1968) *Papers on Time and Tense*, Oxford: Clarendon Press. (Papers relevant to tense logic and the minimalist tensed theory of time.)
- Smith, Q.** (1993) *Language and Time*, New York: Oxford University Press. (An explication and defence of the maximalist tensed theory of time, and criticism of the minimalist tensed theory and the tenseless theory of time.)
- Smith, Q and Oaklander, L.N.** (1995) *Time, Change and Freedom*, London: Routledge. (An introductory discussion of the tensed and tenseless theories of time and their relation to other topics in metaphysics.)

Theory of types

The theory of types was first described by Bertrand Russell in 1908. He was seeking a logical theory that could serve as a framework for mathematics, and, in particular, a theory that would avoid the so-called 'vicious-circle' antinomies, such as his own paradox of the property of those properties that are not properties of themselves - or, similarly, of the class of those classes that are not members of themselves. Such paradoxes can be thought of as resulting when logical distinctions are not made between different types of entities, and, in particular, between different types of properties and relations that might be predicated of entities, such as the distinction between concrete objects and their properties, and the properties of those properties, and so on. In 'ramified' type theory, the hierarchy of properties and relations is, as it were, two-dimensional, where properties and relations are distinguished first by their order, and then by their level within each order. In 'simple' type theory properties and relations are distinguished only by their orders.

1 Ramified type theory

The theory of (ramified) types was first described by Bertrand Russell (1908) as a way of avoiding vicious-circle antinomies, including the one discovered by Russell himself regarding the property that applies to all and only those properties that do not apply to themselves (see [Russell, B.A.W. §§7-8; Paradoxes of set and property §5](#)). A property is said to be 'impredicable' if, and only if, it is not predicable of itself; thus, the property of being impredicable is impredicable if, and only if, it is not impredicable.

$\text{Impred}(\text{Impred}) \equiv \neg \text{Impred}(\text{Impred})$

From this the contradiction follows that the property of being impredicable both is and is not impredicable. The theory of ramified types for propositional functions was developed in *Principia Mathematica* (1910-13). For Russell, such a theory was the framework of a 'logically perfect language', by which he meant a language that would show at a glance the logical structure of the facts that can be described by means of it, and in particular a language 'in which everything that we might wish to say in the way of propositions that are intelligible to us, could be said, and in which, further, structure would always be made explicit' (1959: 165). All that would be needed to add to such a theory is a vocabulary of descriptive constants that correspond to the meaningful words and phrases of the natural sciences or of natural language. The basic laws of a natural science would then be listed as nonlogical postulates of that science, and (as later proposed by Rudolf Carnap) the analytic connections between the words and phrases of natural language would be listed as meaning postulates. The constants and postulates of pure mathematics do not also need to be added, according to Russell, because they are all definable and provable in purely logical terms within the framework of type theory itself - a view known as (Russell's form of) logicism (see [Logicism §1](#)).

The way type theory avoids vicious-circle antinomies is by imposing logico-grammatical constraints on the logic of predicates. The basic idea is that predicates (and the propositional functions they stand for) are to be typed in a hierarchical manner so that predicates of a given type (and their corresponding propositional functions) could never meaningfully be applied to predicates (propositional functions) of the same or higher type. (Thus both $\text{Impred}(\text{Impred})$ and $\neg \text{Impred}(\text{Impred})$ are meaningless in type theory.) In 'ramified' type theory, the hierarchy is, as it were, two-dimensional, having a 'vertical' dimension consisting of a hierarchy of orders and, within each order, a 'horizontal' dimension consisting of a hierarchy of levels. That is, properties are distinguished in ramified type theory as being both of a given order (for example, first-order, second-order and so on) and of a specific level within that order. Thus, for example, where 'being shrewd', 'being courageous', 'being insightful about enemy positions' and so on are examples of first-order/first-level properties that Napoleon and other great generals shared (that is, properties of r-type $(i)/1$, as defined below, assuming Napoleon and other generals to be of r-type i), Napoleon's having *all of the (first-order/first-level) properties of a great general* - that is, all of the (first-order/first-level) properties that make up what it is to be a great general - amounts to Napoleon having a first-order property of level two (that is, a property of r-type $(i)/2$, as defined below). 'Simple' type theory has only the 'vertical' hierarchy.

The description of the ramified hierarchy given by Russell was informal and relied on a use/mention confusion (see [Use/mention distinction and quotation §3](#)) between propositional functions (properties or relations) and the

open formulas with ‘apparent’ (that is, ‘bound’ as opposed to ‘free’) variables taken to represent those functions. The ramified theory was not given a precise, formal description until Alonzo Church’s paper of 1976, which, in accordance with contemporary terminology, uses the notion of ‘order’ to represent the ‘vertical’ hierarchy (instead of Russell’s informal notion of order corresponding to the ‘horizontal’ hierarchy), and which uses the notion of ‘level’ to represent the ‘horizontal’ hierarchy. Church’s definition of ramified types (‘r-types’) is as follows:

- (1) There is an r-type i to which all and only individuals belong, and whose order is stipulated to be 0.
- (2) If m is a non-negative integer and n is a positive integer and $\beta_1, \dots, \beta_m$ are any given r-types, then there is an r-type $(\beta_1, \dots, \beta_m)/n$ to which belong all and only m -ary propositional functions of level n and with arguments of r-types $\leq \beta_1, \dots, \beta_m$, respectively; and the order of such a function is $N + n$, where N is the greatest of the orders corresponding to the types $\beta_1, \dots, \beta_m$ (and $N = 0$ if $m = 0$).

Here, by an ‘individual’, that is, an entity of r-type i , Russell meant a concrete object (in particular, an event) as compared to the propositional functions (abstract properties and relations) that make up the entities of all other r-types. (It is not essential for all logical purposes that the entities of the basic type i be concrete, however.) The notion of the level of a propositional function ϕ of r-type $(\beta_1, \dots, \beta_m)/n$ represents what Russell intended in speaking informally of the ‘apparent’ (bound) variables occurring in ϕ . If N , for example, is the greatest of the orders of $\beta_1, \dots, \beta_m$, and k is the greatest of the orders of the ‘apparent’ variables occurring in ϕ (construed as a formula of r-type $(\beta_1, \dots, \beta_m)/n$), then $n = 1$ if $k \leq N$, and $n = k + 1$ if $N < k$. According to Russell, a propositional function (or formula for such) was said to be ‘predicative’ if ‘it is of the lowest order compatible with its having the arguments it has’ (1910-13: vol. 1, 53); that is, in terms of the notion of level, ϕ is predicative if, and only if, $n = 1$.

In the simple theory of types the distinction between levels is dropped and only the distinction between orders is retained, so that in effect all propositional functions of the simple theory are predicative - although the comprehension principle of simple type theory, unlike that of ramified type theory, allows that such functions can be specified ‘impredicatively’, that is, in terms of a totality to which they belong.

2 The axiom of reducibility

It was Russell’s view that all of traditional mathematics about numbers as abstract objects could be represented in terms of the predicative propositional functions of his theory of ramified types. He could not justify such a view, however, without assuming an ‘axiom (schema) of reducibility’, which stipulates that every propositional function of whatever r-type is coextensive with a predicative propositional function (having arguments of the same r-type). In symbols:

$$(\forall \phi)(\exists \psi)(x_1) \dots (x_m)[\phi(x_1, \dots, x_m) \equiv \psi!(x_1, \dots, x_m)],$$

where ‘ ϕ ’ and ‘ ψ ’ are variables for propositional functions of r-types $(\beta_1, \dots, \beta_m)/n$ and $(\beta_1, \dots, \beta_m)/1$, respectively, and ‘ $x_1, \dots, x_m$ ’ are variables of r-type $\beta_1, \dots, \beta_m$, respectively. (Russell used ‘ $\psi!$ ’ to indicate that the propositional function represented by ‘ ψ ’ is predicative.) This axiom was criticized by a number of logicians (including L. Chwistek 1922; F. Ramsey 1931; W.V. Quine 1936) who argued that, on the assumption that type theory is an extensional system in which propositional functions of the same r-type are identical if they are coextensive (that is, true of all the same objects), the effect of the axiom of reducibility was to nullify the distinction between the different levels of propositional functions - in which case ramified type theory amounted to nothing more than a notationally complex version of simple type theory.

The extensionality assumption is dubious, however, if propositional functions are interpreted intensionally as properties and relations (as Russell originally intended). The point of the axiom of reducibility was not to identify properties and relations with their extensions, but to represent their extensions within the wider intensional framework, which is why Russell also called it ‘the axiom of classes’ (1910-13: vol. 1, 167). The effect of the axiom is that so-called ‘impredicative’ specifications of propositional functions - that is, specifications of functions in terms of a totality to which they belong - are ‘annulled in extensional but not in intensional matters’ (Church 1976: 758).

Ramsey emphasized a distinction between so-called logical and semantic antinomies, both of which the theory of

ramified types was designed to avoid. Russell's paradox, for example, is one of the logical antinomies, as is Georg Cantor's paradox that there are classes (such as the universal class) larger than all other classes, even though, by a theorem of Cantor's, there is a class larger than any given class. These are paradoxes that, without some restrictions such as those imposed in a theory of types, simple or ramified, might occur in any system rich enough to represent properties as objects having properties, or classes as members of classes. The semantic antinomies, on the other hand, assume special semantic notions, such as truth (as in the liar paradox of a sentence that says of itself that it is false, which if true is false, and if false is true), or definability (as in G.G. Berry's paradox of 'the least integer not nameable in fewer than nineteen syllables', which is itself a name with just eighteen syllables), and other related notions of semantics. Ramsey (1931) thought of the semantic paradoxes as 'epistemological' in nature. (See [Paradoxes of set and property](#).)

3 The simple v. the ramified theory of types

The simple theory of types suffices to avoid the logical paradoxes, but not the semantic ones, unless a distinction is made (as first proposed by A. Tarski) between the object language of simple type theory and a metalanguage in which the different semantic notions for the expressions of simple type theory can be defined or developed axiomatically (see [Tarski's definition of truth](#)). The ramified theory of types avoids the semantic paradoxes by incorporating an additional hierarchy of levels that (combined with the hierarchy of orders) implicitly contains a hierarchy of languages, and that in effect amounts to a special form of the metalanguage/object language distinction by which the semantic (and related intensional) paradoxes can be avoided (as noted by [Church 1976: §4](#)). Adding the axiom of reducibility does not nullify the way ramified type theory resolves or avoids the semantic (and related intensional) paradoxes. (See Church (1976: §5) and Myhill (1979) for justification of this claim.)

Different versions of the simple theory of types have been formulated by L. Chwistek (1922), F. Ramsey (1931), R. Carnap (1929), K. Gödel (1931), and A. Church (1940). The simple theory is not committed to an extensional logic, although it is often formulated as such. R.M. Montague (1974), for example, extended Church's simple theory of types by incorporating into it operators for representing the intensions and extensions of expressions, and then used the resulting system as an intensional logic by which to provide an interpretation of a variety of intensional contexts of natural language (along lines originally described by Frege). Church (1974) has also formulated an intensional version of the simple theory by adding to it a formula operator for the Russellian notion of propositional identity. But in 1984, Church also showed that, given Russell's notion of a proposition, and hence of propositional identity, Russell's antinomy of propositions ([Russell 1903: §500](#)) can be reconstructed in simple type theory, but not in ramified type theory. According to this antinomy, there cannot be more classes of propositions than there are propositions - because every non-empty class of propositions can be associated with its 'logical product', which, by Russell's notion of propositional identity, is unique for that class; and yet, by Cantor's theorem, there must be more classes of propositions than there are propositions (see [Cantor's theorem](#)).

See also: [Church, A. §3; Intensional logic; Logical and mathematical terms, glossary of](#)

NINO B. COCCHIARELLA

References and further reading

- Carnap, R.** (1929) *Abriss der Logistik, mit besonderer Berücksichtigung der Relationstheorie und ihrer Anwendungen* (Outline of Logistic, with Special Consideration of the Theory of Relations and their Applications), Vienna: Springer. (This work includes an early formulation of simple type theory and its applications as a logical framework for the axiomatic method. Examples of axiomatic theories, such as set theory, Peano arithmetic, projective geometry and space-time topology are given within the simple type theory formulated. The book was later expanded and developed as Carnap (1954).)
- Carnap, R.** (1954) *Einführung in die symbolische Logik, mit besonderer Berücksichtigung ihrer Anwendungen*, Vienna: Springer; trans. *Introduction to Symbolic Logic and its Applications*, New York: Dover, 1958. (A simple, clear description of Carnap's version of simple type theory can be found in this book, which also includes many important applications of type theory to the construction of the different kinds of numbers, the foundations of mathematics, spacetime physics, anthropological kinship theories, and axiomatic biology.)
- Church, A.** (1940) 'A Formulation of the Simple Theory of Types', *Journal of Symbolic Logic* 5: 56-68. (Elegant, but somewhat difficult for students with only an introductory understanding of logic.)
- Church, A.** (1974) 'Russellian Simple Type Theory', *Proceedings and Addresses of the American Philosophical*

- Association 47: 21-33.(This paper includes a formulation of simple type theory with a primitive operator for propositional identity.)
- Church, A.** (1976) 'Comparison of Russell's Resolution of the Semantical Antinomies with That of Tarski', *Journal of Symbolic Logic* 41: 747-60.(An intermediate level understanding of logic should suffice to read this paper, which includes some important observations on how ramified, but not simple, type theory can avoid certain semantic and intensional paradoxes.)
- Church, A.** (1984) 'Russell's Theory of Identity of Propositions', *Philosophia Naturalis* 21 (2-4): 513-22. (Reviews Russell's antinomy about propositions (Russell 1903: §500). Church shows that this antinomy cannot be resolved in simple type theory, given Russell's view of propositional identity, but it can be resolved in ramified type theory.)
- Chwistek, L.** (1922) 'Über die Antinomien der Prinzipien der Mathematik' (On the Antinomies of the Principles of Mathematics), *Mathematische Zeitschrift* 14: 236-43.(Now read mostly for historical interest.)
- Cocchiarella, N.B.** (1987) *Logical Studies in Early Analytic Philosophy*, Columbus, OH: Ohio State University Press.(Chapter 1 includes a detailed description of the theory of logical types in the development of Russell's views from his 1903 analysis of his paradox to the 1910-13 formulation of his ramified theory. Chapter 2 compares and reconstructs without paradox both Russell's and Frege's early versions of logicism. Chapter 5 explains how Russell's 1925 modifications to his ramified theory commit him to a system too weak for his earlier logicism.)
- Copi, I.** (1971) *The Theory of Logical Types*, London: Routledge & Kegan Paul.(A good, elementary introduction.)
- Gödel, K.** (1931) 'Über formal unentscheidbare Sätze der *Principia Mathematica* und verwandter Systeme I', *Monatshefte für Mathematik und Physik* 38: 173-98; trans. 'On Formally Undecidable Propositions of *Principia Mathematica* and Related Systems', in J. van Heijenoort (ed.) *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard University Press, 1967, 592-617.(A classic, including Gödel's famous incompleteness theorems as applied to a type-theoretical development of arithmetic.)
- Montague, R.** (1974) *Formal Philosophy: Selected Papers of Richard Montague*, ed. R.H. Thomason, New Haven, CT: Yale University Press.(See especially the essays 'Universal Grammar' and 'The Proper Treatment of Quantification in Ordinary English', for Montague's development of Church's simple type theory as an intensional logic.)
- Myhill, J.** (1979) 'A Refutation of an Unjustified Attack on the Axiom of Reducibility', in G.W. Roberts (ed.) *Bertrand Russell Memorial Volume*, London: Allen & Unwin.(Readable with an intermediate level of understanding of logic; explains how the addition of the axiom of reducibility to ramified type theory results in a system essentially different from simple type theory.)
- Quine, W.V.** (1936) 'On the Axiom of Reducibility', *Mind* 45: 498-500.(One of the earliest arguments that the addition of the axiom of reducibility to ramified type theory amounts to a system only notationally different from simple type theory.)
- Ramsey, F.P.** (1931) 'The Foundations of Mathematics', in *The Foundations of Mathematics and Other Logical Essays*, ed. R.B. Braithwaite, London: Routledge & Kegan Paul.(Ramsey's version of simple type theory, and his arguments why, as a foundation for mathematics, the simple theory is to be preferred to the ramified theory.)
- Russell, B.A.W.** (1903) *The Principles of Mathematics*, Cambridge: Cambridge University Press; 2nd edn, London: Allen & Unwin, 1937; repr. London: Routledge, 1992, esp. ch. 10.(Russell's first extended discussion of his paradox, along with some proposals for avoiding it. His first rough ideas about a theory of types are described in appendix B, which also contains an account, in §500, of his antinomy about propositions.)
- Russell, B.A.W.** (1908) 'Mathematical Logic as Based on the Theory of Types', *American Journal Of Mathematics* 30: 222-62; repr. in *Logic and Knowledge, Essays 1901-1950*, ed. R.C. Marsh, London: Allen & Unwin, 1956, 59-102; repr. London: Routledge, 1992.(Accessible to readers with an intermediate-level understanding of logic.)
- Russell, B.A.W.** (1959) *My Philosophical Development*, London: Allen & Unwin; repr. London: Routledge, 1993. (An autobiographical account of the development of Russell's philosophical and logical views.)
- Whitehead, A.N. and Russell, B.A.W.** (1910-13) *Principia Mathematica*, Cambridge: Cambridge University Press, 3 vols; 2nd edn, 1925-7.(Development of the theory of ramified types.)

Turing machines

Turing machines are abstract computing devices, named after Alan Mathison Turing. A Turing machine operates on a potentially infinite tape uniformly divided into squares, and is capable of entering only a finite number of distinct internal configurations. Each square may contain a symbol from a finite alphabet. The machine can scan one square at a time and perform, depending on the content of the scanned square and its own internal configuration, one of the following operations: print or erase a symbol on the scanned square or move on to scan either one of the immediately adjacent squares. These elementary operations are possibly accompanied by a change of internal configuration. Turing argued that the class of functions calculable by means of an algorithmic procedure (a mechanical, stepwise, deterministic procedure) is to be identified with the class of functions computable by Turing machines. The epistemological significance of Turing machines and related mathematical results hinges upon this identification, which later became known as Turing's thesis; an equivalent claim, Church's thesis, had been advanced independently by Alonzo Church. Most crucially, mathematical results stating that certain functions cannot be computed by any Turing machine are interpreted, by Turing's thesis, as establishing absolute limitations of computing agents.

1 Human computers

In his seminal article of 1937, 'On Computable Numbers, With an Application to the *Entscheidungs-problem* [decision problem]', Alan Mathison Turing argued that a function is calculable by some algorithmic procedure if and only if it is computable by a Turing machine. The main claim in Turing's argument is that the behaviour of a human being carrying out an algorithmic procedure (adopting the terminology of Gandy (1988), a 'computer') is constrained by limitations of human memory and sensory capacities. It is appropriate to review briefly these restrictive conditions, for they are reflected in the mathematical definition of a Turing machine. (For a careful analysis of Turing's argument, see Sieg (1994); an analogue of Turing's argument for parallel computations by digital computers is given by Gandy (1980).)

Turing considered a 'one-dimensional' computing space formed by a potentially infinite tape uniformly divided into squares, in each of which at most one symbol may be written. What are the possible processes which can be carried out by a computer working on this computing space?

- (1) The computer can write on the tape symbols from a fixed finite list (for example, letters of the alphabet). This restriction is argued for on the ground that if one allowed an infinity of symbols, then one would have symbols differing to an arbitrarily small extent (which, presumably, could not be unambiguously recognized or written down by the computer in view of sensory and memory limitations).
- (2) The computer can observe squares on the tape. Limitations of human visual apparatus seem sufficient to justify the existence of a fixed finite bound B on the number of contiguous squares observable at any one moment.
- (3) The computer can remember earlier stages of the computation and bring this information to bear on future behaviour. This memory consists in what Turing calls the computer's 'states of mind'. There is a fixed finite bound on the number of 'states of mind' which need be taken into account in a mechanical calculation. If one admitted the possibility of using infinitely many states of mind, Turing observes, some of them would be arbitrarily close and would be confused by the computer.
- (4) The elementary actions the computer can perform consist of a change of either (a) observed squares or (b) the contents of some squares, possibly accompanied by a change of 'state of mind'. Any other operation of the computer can be viewed as a composition of these elementary actions. Furthermore, Turing assumes that (c) each of the newly observed squares must be within a bounded distance L of an immediately previously observed square, and that nothing will be lost by supposing that only the contents of observed squares can be changed.
- (5) Which elementary action the computer is to undertake at a given moment depends only on what the computer currently sees and remembers. Thus, the observed squares and the 'state of mind' at any moment uniquely determine the action to be undertaken.

These restrictive conditions ensure that the computer's behaviour is fixed by a finite list of instructions: there are only finitely many different types of observable configurations of symbols (by (1) and (2)) and finitely many 'states of mind' (by (3)); to each pair formed by a 'state of mind' and type of observable configuration is uniquely

associated (by (5)) an elementary action (fixed by (4)).

2 (Universal) Turing machines

Turing claims that the behaviour of a computer satisfying the restrictive conditions above can be faithfully simulated by a machine organized as follows. The internal configurations $q_1, \dots, q_n$ of the machine correspond to the states of mind $m_1, \dots, m_n$ of the computer. The machine scans B squares corresponding to the B squares observed by the computer. In any of its actions, the machine can change a symbol on one of the scanned squares or shift attention to observe another configuration of symbols in such a way that condition (4)(c) is satisfied. The action undertaken at any moment by the machine, and its next internal configuration, are uniquely determined by the scanned symbols and the internal configuration at that moment. In turn, one can show that any computation of this machine can be carried out by a machine M endowed with a more elementary repertoire of actions, which provides an intuitive visualization of a Turing machine. This machine has a fixed finite number of internal configurations; it is capable of scanning only one square at a time (that is, $B = 1$); it may print or erase a symbol on the scanned square or else move on to scan either of the immediately adjacent squares (that is, $L = 1$).

To the machine M one can associate a function $f(x_1, \dots, x_n)$ of n non-negative integers in the following way. Given an n -tuple $a_1, \dots, a_n$, since the machine can print at least one symbol, say 1, the input $\langle a_1, \dots, a_n \rangle$ can be represented on an otherwise blank tape by a sequence of $a_1 + 1$ consecutive 1's followed by a blank square followed by a sequence of $a_2 + 1$ consecutive 1's followed by a blank square... followed by a sequence of $a_n + 1$ consecutive 1's. The machine is to be started in a specific internal configuration, q_1 , scanning the leftmost square containing a 1. The output number (if any), which is to be identified with $f(a_1, \dots, a_n)$, is given by the number of 1's written on the tape to the right of the scanned square when (and if) the machine halts. These stipulations enable one to associate to M an n -ary *partial* (rather than total) function (that is, a function which may be undefined for some n -tuple of non-negative integers), because for some input the machine may fail to provide an output by not halting.

The initial stage of computation of M on the input $\langle a_1, \dots, a_n \rangle$ has been described by giving (1) the tape content, (2) the machine's internal configuration, and (3) the scanned square. Each subsequent stage of the computation can be described by providing this information, which is called an 'instantaneous description' of the machine. Any computation which produces a final output can be described by a finite sequence of instantaneous descriptions.

The essential features of this type of machine can be captured by precise mathematical definitions. Let us consider the list of symbols $S_0, S_1, S_2, \dots, q_1, q_2, q_3, \dots, R$ and L , and sequences of symbols from this list having one of the following forms:

- (1) $q_i S_j S_k q_l$
- (2) $q_i S_j R q_l$
- (3) $q_i S_j L q_l$.

Sequences of this form are called 'quadruples'. Turing machines will be now defined (following Davis 1958) entirely in terms of quadruples, which can be interpreted as specifying the next action of a Turing machine when its internal configuration is q_i and the symbol it is scanning is S_j . Quadruple (1) indicates that S_j is to be replaced by S_k and the machine is to enter internal configuration q_l ; quadruples (2) and (3) indicate that the machine is to shift one square to the right or to the left, respectively, and then enter internal configuration q_l . We stipulate that the symbol S_0 will serve as a blank; thus, replacing symbol S_j by S_0 amounts to erasing it.

Now a Turing machine is a finite, non-empty set of quadruples that contains no two quadruples with the same first two symbols. The latter condition guarantees the deterministic character of the machine operation, since no more than one instruction will be applicable at any given moment.

This definition identifies a Turing machine with a finite table of instructions in a certain normal form. Equally rigorous definitions can be given for the notions of instantaneous description and computation, as well as for the Turing computable functions, that is, the class of partial functions that Turing machines can compute (see Davis 1958: ch. 1). A basic result of computability theory is the extensional equivalence between this class of functions and the partial recursive functions. Analogous results have been obtained for classes of functions defined in terms of other algorithmic processes (Markov algorithms, lambda-calculi, Post's production systems, Kleene's

equational calculus, and so on; see [Church's thesis](#); [Computability theory](#)).

One can define an algorithmic one-one correspondence of the members of the class of Turing machines with the set of all non-negative integers. This fact entails that there is an algorithm (and, by Turing's thesis, a Turing machine) enabling one to retrieve, given any integer x , the Turing machine indexed by x and, conversely, given any Turing machine, the integer associated with it (see [Rogers 1958](#)). This mapping yields an enumeration $M_0, M_1, M_2, \dots$ of the class of Turing machines. Correspondingly, we shall indicate by f_x the partial function computable by M_x .

Under any such enumeration, there is a number u such that for any numbers x and y , the machine M_u computes the following binary function:

$$f_u(x, y) = \begin{cases} f_x(y) & \text{if } f_x(y) \text{ is defined} \\ \text{undefined} & \text{otherwise.} \end{cases}$$

Thus, the machine M_u can be employed to compute any unary Turing computable function. Furthermore, M_u can also compute the n -ary Turing computable functions, provided one codes any n -tuple $\langle y_1, \dots, y_n \rangle$ of input numbers by means of a non-negative integer. Again, this coding can be given by means of Turing computable functions (see [Davis 1958](#): ch. 4). In view of these properties, M_u is called a universal Turing machine. A universal Turing machine is explicitly described in Turing ([1937](#)).

3 The horizon of computability (and beyond)

There are number-theoretic functions which cannot be computed by any Turing machine. The mere existence of such functions follows from the fact that the class of Turing computable functions is denumerable (it can be put into a one-one correspondence with the non-negative integers) whereas, by Cantor's theorem (see [Cantor, G.](#)), there are indenumerably many number-theoretic functions.

Specific examples of non-computable functions were provided by Turing in his 1937 paper. Consider the so-called 'halting problem': is there a Turing machine capable of deciding for arbitrary x and y whether the Turing machine M_x will eventually halt after being started on input y ? In other words, is the total number-theoretic function

$$h(x, y) = \begin{cases} 1 & \text{if } f_x(y) \text{ is defined} \\ 0 & \text{otherwise} \end{cases}$$

a Turing computable function? The answer to this question is negative: the function $h(x, y)$ is not Turing computable, and the halting problem is said to be unsolvable. Turing used the unsolvability of the halting problem to establish a negative solution to Hilbert's *Entscheidungsproblem* [decision problem] (see [Church's theorem and the decision problem](#); [Hilbert's programme and formalism](#)).

The unsolvability of the halting problem implies that the behaviour of some Turing machines cannot be completely predicted by any Turing machine. An example of such 'unpredictable' machines is just the universal Turing machine M_u . The existence of a Turing machine capable of determining, for every x and y , whether the machine M_u will eventually halt when started on input $\langle x, y \rangle$ conflicts with the negative solution to the halting problem.

The import of this and other 'negative' results concerning Turing machines crucially hinges upon Turing's thesis. Indeed, if a function cannot be computed by a Turing machine then, by Turing's thesis, one can infer that there is no algorithm enabling one to calculate the values of this function. In particular, no algorithmic procedure can enable one to answer, for each x and y , all questions of the form 'Does M_x halt on y ?' or, equivalently, 'Does x belong to the set K_y formed by the numbers z such that $f_z(y)$ is defined?'.

Suppose we are provided with some unspecified means - an oracle, as it were - for solving problems of the form 'Does x belong to X ?', where X is a set of non-negative integers. One may then define a new class of machines, which one may call O -machines, suitably extending the definition of Turing machine so as to allow for quadruples of the form

$$(4) \quad q_i S_j q_k q_l,$$

and associating to each *O*-machine a specific set X of non-negative integers. An *O*-machine in internal configuration q_i which is scanning symbol S_j may be interpreted as asking the oracle: ‘Does the number x of occurrences of 1 on my tape at this moment belong to X ?’ If the answer is ‘yes’ then the *O*-machine changes its internal configuration to q_k ; if the answer is ‘no’ then the *O*-machine changes its internal configuration to q_l (see [Davis 1958: ch. 1](#)). The function computable by an *O*-machine is said to be computable in the set X associated with that machine.

A class of functions computable in a set X may include functions that are not Turing computable. Furthermore, given a set X , one can prove the existence of functions that are not computable in X , that is, functions which cannot be computed by *O*-machines having access to an oracle capable of deciding membership of the set X . Results of this kind are central to the theory of the degrees of unsolvability, which introduces a hierarchy of classes of unsolvable problems.

One may also distinguish between classes of solvable problems, in terms of the amount of resources a computing machine may need for their solution. Among the various, equivalent models that are available, the model of computation commonly chosen to estimate these resources is that of the Turing machine (see [Complexity, computational](#)).

The notion of (universal) Turing machine also plays a prominent role in philosophical discussions of the mind-machine problem. Theoretical limitations of (universal) Turing machines have been appealed to in arguments purporting to show that the human mind is not subject to analogous limitations. Turing ([1950](#)) employed this notion in order to circumscribe the class of devices that are admitted to compete with human beings in his test for intelligent behaviour. More recently, (universal) Turing machines have been used to isolate distinguishing features of the notion of computational explanation in cognitive science, as well as heuristic ideas underlying research programmes in artificial intelligence (see, for example, [Newell and Simon 1976](#)).

See also: [Logical and mathematical terms, glossary of](#)

GUGLIELMO TAMBURRINI

References and further reading

- Davis, M.** (1958) *Computability and Unsolvability*, New York: McGraw-Hill; repr. New York: Dover, 1982. (Classic introduction to Turing machines and the theory of computability. The first part of this book is accessible to readers with no special mathematical training.)
- Davis, M.** (ed.) (1965) *The Undecidable: Basic Papers on Undecidable Propositions, Unsolvability Problems and Computable Functions*, Hewlett, NY: Raven Press. (An anthology of fundamental papers on undecidability and unsolvability, which includes [Turing \(1937\)](#).)
- Gandy, R.** (1980) ‘Church’s Thesis and Principles for Mechanisms’, in J. Barwise, H.J. Keisler and K. Kunen (eds) *The Kleene Symposium*, Amsterdam: North Holland, 123-45. (Presents an analogue of Turing’s argument for parallel computations by digital computers.)
- Gandy, R.** (1988) ‘The Confluence of Ideas in 1936’, in R. Herken (ed.) *The Universal Turing Machine: A Half-Century Survey*, Oxford: Oxford University Press, 55-112. (An informative discussion of the conceptual background of Turing’s analysis.)
- Herken, R.** (ed.) (1988) *The Universal Turing Machine: A Half-Century Survey*, Oxford: Oxford University Press. (See especially Kleene’s and Gandy’s articles. Feferman’s article reviews the ideas leading to the notion of oracle machine.)
- Kleene, S.C.** (1988) ‘Turing’s Analysis of Computability, and Major Applications of It’, in R. Herken (ed.) *The Universal Turing Machine: A Half-Century Survey*, Oxford: Oxford University Press, 17-54. (Includes a terse outline of Gödel’s incompleteness theorems, highlighting their relation to Turing machines in a form accessible to non-specialists.)
- Newell, A. and Simon, H.** (1976) ‘Computer Science as Empirical Enquiry: Symbols and Search’, *Communications of the ACM* 19: 113-26. (Formulates broad heuristic hypotheses in artificial intelligence and cognitive science by crucial appeal to universal computing machines.)
- Rogers, H.** (1958) ‘Gödel Numbers of Partial Recursive Functions’, *Journal of Symbolic Logic* 23: 331-41. (A technical analysis of the numerical codings referred to in §2.)
- Sieg, W.** (1994) ‘Mechanical Procedures and Mathematical Experience’, in A. George (ed.) *Mathematics and Mind*, Oxford: Oxford University Press, 71-117. (Includes a careful analysis of Turing’s argument.)

- Turing, A.M.** (1937) 'On Computable Numbers, With an Application to the *Entscheidungsproblem* [decision problem]', *Proceedings of the London Mathematical Society*, series 2, 42: 230-65; repr. in M. Davis (ed.) *The Undecidable*, Hewlett, NY: Raven Press, 1965. (Turing's seminal article in the theory of computability: §9 presents the major argument for Turing's thesis.)
- Turing, A.M.** (1950) 'Computing Machinery and Intelligence', *Mind* 59: 433-60; repr. in D.C. Ince (ed.) *Collected Works of A.M. Turing*, vol. 3, *Mechanical Intelligence*, Amsterdam: North Holland, 1992. (An influential article on the mind-machine problem, presenting Turing's reflections on (universal) computing machines and intelligent behaviour.)

Turing reducibility and Turing degrees

A reducibility is a relation of comparative computational complexity (which can be made precise in various non-equivalent ways) between mathematical objects of appropriate sorts. Much of recursion theory concerns such relations, initially between sets of natural numbers (in so-called classical recursion theory), but later between sets of other sorts (in so-called generalized recursion theory). This article considers only the classical setting. Also Turing first defined such a relation, now called Turing- (or just T-) reducibility; probably most logicians regard it as the most important such relation. Turing- (or T-) degrees are the units of computational complexity when comparative complexity is taken to be T-reducibility.

1 Usage and notation

ω = the set of all natural numbers; a real is a subset of ω . For sets A, B : $A \cup B = \{x: x \in A \text{ or } x \in B\}$; $A \times B = \{\langle x, y \rangle: x \in A, y \in B\}$; for $n \in \omega$, $n \neq 0$: $A^n = A \times \dots \times A$ (taking A n -times). For a function φ : $\text{dom}(\varphi)$ = the domain of φ ; φ is a partial function on ω^n iff $\text{dom}(\varphi) \subseteq \omega^n$; such a function is total iff its domain is ω^n . In this article a partial function will be partial on ω^n for some $n \in \omega$; φ is total iff $\text{dom}(\varphi) = \omega^n$. (So a total function on ω^n is also partial - an unfortunate, but standard, usage.) For a real X , χ_X , is X 's characteristic function: $\chi_X(x) = 1$ if $x \in X$; $\chi_X(x) = 0$ if $x \notin X$.

2 Informal characterization of T-reducibility

For reals X and Y : X is reducible to Y iff Y is at least as computationally complex (in a sense whose specification determines the reducibility relation) as X is, that is, given any $n \in \omega$ we could answer the question ' $n \in x$?' by means that are in some sense computational except (perhaps) for using (subject to various possible constraints) answers to membership-questions about Y . For T-reducibility, we require computation to be effective and impose no restraints. A relativized algorithm consists of instructions like an 'absolute' algorithm (see [Computability theory](#)) except that they may tell one to 'ask a question of an oracle', one of the form 'what about y ?', where y can be any natural number given to the oracle as a numeral; running the algorithm relative to Y , in a finite time the oracle says 'yes' if $y \in Y$, and 'no' otherwise. Suppose A is such a relativized algorithm, and we have an oracle for the real Y and infinite amount of time. Were we to apply A to a given input, we would grind along following A as we would an 'absolute' algorithm, except that from time to time A might specify a natural number (in some sort of numeral form) for us to ask the oracle about; the oracle would answer; we would grind on following A , exactly what A requires us to do being 'mechanically' determined by the answers the oracle gave to this and previous queries; eventually A might tell us to stop, delivering an output; A also might leave us computing forever. The oracle is oracular in that we do not trouble ourselves about how it comes up with its answers; going from a query of the oracle to its answer is a non-effective step in an otherwise effective computation. Now we will introduce some 'informal' definitions.

(I) A partial function φ on ω is computed by an algorithm A relative to a real Y iff for any $x \in \omega$, were we to run A on input x using an oracle for Y : if $x \in \text{dom}(\varphi)$ the computation would halt giving the value $\varphi(x)$; if $x \notin \text{dom}(\varphi)$ the computation would not halt. φ is computable relative to Y iff φ is computed by some such A relative to Y . X is decidable relative to Y iff χ_X is computable relative to Y (that is, for some A for any $x \in \omega$, were we to run A on input x using an oracle for Y , the computation would eventually halt with a correct answer (1 = 'yes', 0 = 'no') to the question ' $x \in X$?'. X is recursively enumerable relative to Y iff X is the range of a function that is total on ω and computable relative to Y .

3 Two rigorous definitions

One way to replace these informal definitions by rigorous definitions uses the notion of an oracle machine (that is, a relativized Turing machine (see [Turing machines](#))), credited by Rogers (1967) to M. Davis. Fix a finite set ALPH (an alphabet) whose members we will call symbols, say with 0 (= blank), $1, 2 \in \text{ALPH}$; without mathematical loss, we could suppose that $\text{ALPH} = \{0, 1, 2\}$. Recall the three usual rule-schemata defining being a Turing-machine table (for a single tape infinite in both directions and with ' 0 ' meaning 'blank') relative to ALPH and to a set STATE of internal states. Add to them this fourth schema (Query): for y = the number of occurrences of 1 on the tape at the moment, if in state q then: if the oracle says 'yes' to y then change to state q' ; if it says 'no'

change the state q'' . (By using occurrences of 2 rather than 1, the device can record and preserve results of previous steps while ‘asking the oracle a question’, insulating this record from the question asked.)

Pick $L, R \notin \text{ALPH}$, $\text{ACT} = \text{ALPH} \cup \{L, R\}$. Represent any Turing-machine rule by a quadruple of the form $\langle q, s, a, q' \rangle$ for $q \in \text{STATE}$, $s \in \text{ALPH}$, $a \in \text{ACT}$, $q' \in \text{STATE}$, with this understanding: if $a \in \text{ALPH}$ the instruction says ‘Erase the cell and write a ’; L and R represent ‘Move the tape one cell to the left (that is, the head one cell to the right)’ and ‘Move the tape one cell to the right’, respectively. Any rule of form (Query) can be represented by a quadruple $\langle q, s, a, q'q'' \rangle$ for $q, q', q'' \in \text{STATE}$, $s \in \text{ALPH}$.

Let an oracle-machine table on $\langle \text{ALPH}, \text{STATE} \rangle$ be a function from $\text{STATE} \times \text{ALPH}$ into $(\text{ACT} \cup \text{STATE}) \times \text{STATE}$. Identifying a function f with the set of ordered pairs of the form $\langle x, f(x) \rangle$, and identifying an ordered pair of the form $\langle \langle a, b \rangle, \langle cd \rangle \rangle$ with $\langle a, b, c, d \rangle$, such a table is a set of quadruples coding instructions. For mathematical purposes, an oracle machine is an ordered pair $M = \langle T, q_0 \rangle$ for T an oracle-machine table and $q_0 \in \text{STATE}$; q_0 is M ’s initial state. Given a tape with cells in any condition, a particular cell on that tape, and a real Y , if each query of the oracle is true to Y (that is, if the device is in state q reading s , y is as above, and $T(\langle q, s \rangle) = \langle q', q'' \rangle$, then the machine goes into state q' if $y \in Y$, and goes into state q'' otherwise), an oracle machine M carries out a uniquely determined succession of operations starting in state q_0 with its head on that chosen cell, one which may or may not terminate in a finite number of steps; this is its computation (also known as its run) relative to Y . The definition of an oracle machine does not involve a particular value for ‘ Y ’: a given oracle machine M with a given input can run with any real Y , perhaps producing different computations for different reals. Also, talk of tapes and devices is heuristic, replaceable by an austere but tedious definition of the computation M carries out input x using Y .

(II) We associate with each such M and Y a partial function φ on ω as follows: represent a natural number x as input by string of $x + 1$ consecutive occurrences of 1 on the tape with all other cells blank; adding 1 distinguishes input 0 from starting with a blank tape. Start M in state q_0 reading the leftmost cell containing a 1; as output take the total number of occurrences of 1 appearing anywhere on the tape when and if the machine stops; thus $\varphi(x) =$ that total number of 1s; $x \in \text{dom}(\varphi)$ iff M eventually halts when run with input x ; if M does not halt, $x \notin \text{dom}(\varphi)$ and $\varphi(x)$ does not exist (as mathematicians say, ‘ $\varphi(x)$ is undefined’). This φ is the partial function computed by M using Y . φ is T-computable from Y iff φ is computed by some oracle machine using Y . X is T-reducible to Y (in symbols $X \leq_T Y$) iff χ_X is T-computable from Y .

For $\text{ALPH} = \{0, 1, 2\}$, represent the possible ‘actions’ by natural numbers: $i \leq 2$ for ‘Erase the cell and write i ’; 3 for L ; 4 for R ; represent the internal states as natural numbers greater than 4. Represent (or better, define) a machine-table as a function from a finite subset of $\omega \times \{0, 1, 2\}$ into a finite subset of $\omega \times \omega$; such a function can then be represented by a single natural number. An oracle machine M can be coded by a single natural number, M ’s Gödel-number (see **Gödel’s theorems**). For $e = M$ ’s Gödel-number, let φ_Y^e (alternative notation: $\{e\}^Y$) be the partial function M computes using Y .

(III) Turing himself did not use the notion of an oracle machine in his original definition of T-reducibility; rather he defined the class of partial functions (each on ω^n for any $n \in \omega$) recursive in a real Y (also known as the class of functions partial recursive in Y) to be the least class $\underline{C}$ meeting these conditions: (1) all initial functions belong to $\underline{C}$; these are successor, the function mapping each ω^n into $\{0\}$, and all projections (that is, functions of the form $f(x_1, \dots, x_n) = x_i$ for some i , $1 \leq i \leq n$); (2) $\chi_Y \in \underline{C}$; (3) $\underline{C}$ is closed under composition, primitive recursion and minimization. A real X is recursive in Y iff χ_X is.

It is tedious but not hard to prove that this class is exactly the class of partial functions T-computable from Y . Thus X is recursive in Y iff X is T-reducible to Y . In fact, each of the various ways of defining being a recursive function extends naturally to a definition of being a function recursive in an arbitrary real Y . Some regard the coextensiveness of (II), (III), and those other definitions, as evidence that they all rigorously define what (I) informally defines, that is, as evidence for the relativized version of Church’s thesis: a function is computable relative to Y iff it is T-computable relative to Y (see **Church’s thesis**).

4 More definitions and facts

Let $X <_T Y$ iff $X \leq_T Y$ and not $Y \leq_T X$; X and Y are T-equivalent (in symbols $X \equiv_T Y$) iff $X \leq_T Y$ and

$Y \leq_T X$. As the notation suggests: $\leq_T$ is a reflexive transitive relation; $\equiv_T$ is an equivalence relation. We could prove these facts from the mathematical definition of (II) or (III); but it is easier to work informally.

Sample proof Let A and B be algorithms computing χ_X and χ_Y relative to Y and Z respectively. Consider this algorithm: given any x , run A on x ; whenever A tells us to ask the oracle for Y about a y , run B on y , using the oracle for Z; eventually such a run of B halts; if it yields 1 or 0 take that as a ‘yes’ or ‘no’ respectively from an oracle for Y. A halts on an input if responses to its queries of an oracle are true to Y; a run of B using an oracle for Z does deliver the truth about Y (since it computes χ_Y); so this process halts, yielding $\chi_X(x)$, and it uses an oracle for Z (rather than Y, unless $Y = Z$). Thus $\leq_T$ is transitive. (One of recursion theory’s advantages is that we can almost always work in this informal way.)

For a real X let $\deg(x) = \{Y : X \equiv_T Y\}$. Let a be a T-degree iff $a = \deg(x)$ for some real X ; $\mathcal{D}$ = the set of T-degrees. For $a, b \in \mathcal{D}$ let $a \leq b$ iff for some $X \in a, Y \in b, X \leq_T Y$. Note that if $a \leq b$ then for any $X \in a, Y \in b: X \leq_T Y$. The relation $\leq$ is partial ordering, that is, it is reflexive, transitive, and antisymmetric (that is, if $a \leq b$ and $b \leq a$ then $a = b$). Let 0 = the set of recursive reals. For $X, Y \in 0: X \equiv_T Y$; for any real Z if $Z \leq_T X$ then $Z \in 0$; thus $0 \in \mathcal{D}$. T-degrees are sometimes called ‘degrees of unsolvability’, rather misleading since one of them, 0 , represents solvability (= recursiveness). For any real Y and any $X \in 0, X \leq_T Y$; thus 0 is the least T-degree under this (the standard) ordering. Study of the algebraic properties of the structure $\langle \mathcal{D}, \leq \rangle$ is a major component of recursion theory. One basic fact is that the partial ordering $\leq$ is very partial; for $a, b \in \mathcal{D}$, a and b are incomparable iff neither $a \leq b$ nor $b \leq a$; each T-degree is incomparable with continuum-many other T-degrees - as many as there are reals.

Let a degree be recursively enumerable (hereafter r.e.) iff it contains an r.e. real. Recursive reals are r.e.; so 0 is r.e. Let $K^X = \{x : \varphi_x^X(x) \text{ is defined}\}$. In 1944 E. Post gave the first example of a non-recursive r.e. set: for $K = K^{\{\}}$ for $\{\}$ = the empty set; so $\deg(K)$ is another r.e. degree. He went on to pose ‘Post’s problem’: are there other r.e. T-degrees? In 1956 R. Friedberg and A. Muchnik independently proved that there are others; some consider this the first deep result in degree theory. (See [Friedberg 1957](#) or any of the textbooks referred to below.) The study of r.e. degrees is a flourishing corner of recursion theory.

The jump function on the reals assigns each real X to K^X . The effective version of Cantor’s diagonal argument for the uncountability of the set of reals (see [Cantor’s theorem](#)) shows that $K^X \leq_T X$; since $X \leq_T K^X, X \leq_T K^X$. If $X \equiv_T Y$ then $K^X \equiv_T K^Y$; so it makes sense to define the jump function on the T-degrees: the jump of a (usually written a') is the T-degree of K^X for some (thus any) $X \in a$. Thus $a < a'$, showing the jump for degrees is a natural uniform way of transforming a T-degree into a more complex one. Obviously jumping iterates any finite number of times, yielding degrees of increasingly complex sets; in fact, it iterates through the so-called constructibly countable ordinals (see [Hodes 1980](#)), a matter closely connected to Gödel’s constructible hierarchy (see [The constructible universe](#)).

A sentence in a formal language can be effectively coded as a natural number by ‘Gödel numbering’ (see [Gödel’s theorems](#)); so a set of sentences, for example, a theory (in the logician’s sense (see [Model theory](#))), can be coded as a real; so its complexity can be identified with the T-degree of that real. Examples are: the theory of a dense linear ordering without endpoints has degree 0 ; the set of validities of first-order logic is r.e. and so has degree $0'$; ditto for the set of theorems of any axiomatizable theory in a countable language; the set of truths of first-order arithmetic has degree $= 0^\omega =$ the result of jumping $0'$ -times.

S. Simpson (1977) proved that the first-order theory of $\langle \mathcal{D}, \leq, ' \rangle$ (that is, the set of true first-order sentences based on a two-place predicate for $\leq$, a one-place function-constant for the jump, and ‘first-order logic’ with variables ranging over $\mathcal{D}$) was very complicated - of the same T-degree as the set of truths of second-order arithmetic. In 1990 B. Cooper announced a proof (still unpublished at the end of 1993) that the jump-function on $\mathcal{D}$ is first-order definable over $\langle \mathcal{D}, \leq \rangle$, that is, there is a formula $F(x, y)$ with exactly free variables x and y , based on a two place predicate for $\leq$ and ‘predicate logic’, so that with quantifiers ranging over $\mathcal{D}$: for any $a, b \in \mathcal{D}$ $F(a, b)$ is true iff $b = a'$. With Simpson’s theorem, this showed that the first-order theory of $\langle \mathcal{D}, \leq \rangle$ is as complex as the truth-set of second-order arithmetic. Analysis of T-degree of the truth-set have been carried out for many other kinds of structures, including various pieces of $\langle \mathcal{D}, \leq \rangle$.

See also: [Logical and mathematical terms, glossary of](#)

References and further reading

- Friedberg, R.** (1957) 'Two recursively enumerable sets of incomparable degrees of unsolvability', *Proceedings of the National Academy of Science* 43: 236-8. (One of the first two answers to Post's problem.)
- Hodes, H.** (1980) 'Jumping through the transfinite', *Journal of Symbolic Logic* 45: 204-20. (On iterating the jump through the constructibly countable ordinals.)
- Griffor, E.** ed. (forthcoming) *Handbook of Recursion Theory*, Amsterdam: North-Holland Publishing. (Survey articles that will be up-to-date as of publication.)
- Lerman, M.** (1983) *The Degrees of Unsolvability*, Berlin: Springer-Verlag. (For advanced students.)
- Nerode, A., and Shore, R.** (1980) 'Second-order logic and first-order theories of reducibility orderings' in Barwise *et al.* (eds) *Kleene Symposium*, Amsterdam: North-Holland Publishing, 181-200. (A proof of Simpson's theorem, one easier than the original proof.)
- Odifreddi, P.** (1989) *Classical Recursion Theory*, Amsterdam: North-Holland Publishing, Elsevier Science Publishers. (A very good introductory text, emphasising breadth, up-to-date as of 1989; the author plans to extend this with at least one further volume.)
- Rogers, H.** (1967) *The Theory of Recursive Functions and Effective Computability*, New York: McGraw Hill Inc. (The original text on recursion theory, and excellent complement to Odifreddi's text, with very good exercises.)
- Shoenfield, J.** (1971) *Degrees of Unsolvability*, Amsterdam: North-Holland Publishing, and New York: American Elsevier. (Lecture notes, not comprehensive, but very readable.)
- Simpson, S.** (1977) 'First-order theory of the degrees of recursive unsolvability', *Annals of Mathematics* 105: 212-138. (The proof in **Nerode and Shore (1980)** is easier.)
- Soare, R.** (1987) *Recursively Enumerable Sets and Degrees*, Berlin: Springer-Verlag. (For advanced students.)

Type/token distinction

The type/token distinction is related to that between universals and particulars. C.S. Peirce introduced the terms 'type' and 'token', and illustrated the distinction by pointing to two senses of 'word': in one, there is only one word 'the' in the English language; in the other, there are numerous words 'the' on the physical page you are now looking at. The latter are spatiotemporal objects composed of ink; they are said to be word tokens of the former, which is said to be the word type and is abstract. Phonemes, letters and sentences also come in types and tokens.

The distinction between 'type' and 'token', or something very like it, seems to be applicable beyond language to, for example, Beethoven's *Fifth Symphony* and performances of it, the grizzly bear and specimens of it, the Kentucky Derby and runnings of it, and the bubonic plague and outbreaks of it. The type/token distinction is important to linguistics, logic, aesthetics and philosophy of science. In philosophy of mind it is critical to distinguish types of events/states/processes from tokens of them, because those who identify mental events with physical events agree that every mental event token is a physical event token, but divide over whether mental event types, for example, pain, are identical to physical event types, for example, C-fibre stimulation (see [Mind, identity theory of](#)). Just what types are, whether they exist, whether they are 'present in' their tokens and whether all types have tokens are matters of controversy.

To see that we often refer to types, consider the grizzly bear. At one time its US range was most of the west and it numbered 10,000 in California alone. At the end of the twentieth century its range is Montana, Wyoming and Idaho, and it numbers fewer than 1,000. But no particular flesh and blood bear numbers 1,000 or once had a range comprising most of the western USA; if anything, it is a type which does. Similarly, Old Glory had twenty-nine stars in 1846 but fifty in 1996, whether or not any particular American flag underwent such a transformation; Old Glory is also a type.

This last example points to a related notion, that of an 'occurrence'. The stars on Old Glory number fifty; but fifty star *types* or fifty star *tokens*? In fact it cannot be either: not star types, because all the stars on the flag are of the same (five-pointed) type; nor star tokens, because tokens are concrete and the flag in question is abstract and cannot contain concrete parts. Old Glory contains fifty occurrences of the (five-pointed) star. Similarly, the letter 'x' (the very same letter type) occurs three times in the formula (type) ' $(\forall x)(Fx \rightarrow Gx)$ '. Thus the notion of an occurrence of 'x' must not be confused with the notion of a token of 'x' (although it often is so confused). The notion of 'an occurrence of x in y' involves not only x and y, but also how x is situated in y. If we think of a formula as a sequence, then the air of mystery over how the same identical thing can occur twice vanishes. Even concrete particulars can occur more than once (for example, the same person occurs twice in the sequence of New Jersey million dollar lottery winners).

Tokens are particulars. Are types universals? They are if having instances makes something a universal. But in many ordinary and theoretical contexts, terms which refer to types function not as predicates but as singular terms, in sentences that permit existential generalization. ('The grizzly bear ranged over most of the west' entails 'Something ranged over most of the west', for example.) As values of bound first-order variables, they meet a Quinean necessary condition for existence, and thus may also be viewed as particulars (see [Ontological commitment §1](#)).

Do types exist? If types are universals, then the debate over whether they exist goes back at least to Plato. Still, as the preceding paragraph noted, types seem to have an existential advantage over more traditional universals. (Quine held that expression types exist, though not more traditional universals (1987).) Nominalists who deny that there are any abstract objects argue that this is an illusion; that talk about types is just shorthand for talk about tokens. The matter cannot be resolved here, but it may be useful to examine one popular nominalistic account, namely, that every reference to a type (for example, 'The grizzly bear is ferocious') can be replaced without change of information by quantification over all tokens (for example, 'Every grizzly bear is ferocious'). One problem for this account is the existence of numerous counterexamples. 'The grizzly bear ranged over most of the west', for example, cannot be analysed as 'Every grizzly bear ranged over most of the west'. (Perhaps it can be analysed in some other 'type-free' way, but the point is that the nominalistic account under consideration cannot be the systematic one needed to eliminate all the many references to types which occur in our ordinary and

scientific talk.)

A second problem is created by the implication that all types have tokens. Peirce (1931-58) claimed they do, but many thinkers since have denied this. Chomsky (1957), for example, claimed that there are sentences that never have been or will be instantiated - infinitely many. A third problem arises when we try to find a replacement for even a simple true sentence about, say, the word 'one', given the myriad senses it has, and the myriad forms its tokens may take: printed in ink in any number of fonts and handwritings, raised in Braille, incised in marble, existing briefly as so many pixels of light on a computer screen, electronic strings of dots and dashes, smoke signals, hand signals, individual air disturbances produced by human vocal cords, electromagnetic pulses on phone lines, pronounced (and mispronounced) in countless ways in countless accents. Even an (appropriately surrounded) empty space can be a token of 'one'. About the only thing all and only tokens of 'one' clearly have in common is being tokens of 'one' - the type, that is; if that word (type) does not exist, then they would seem to have nothing in common and hence no appropriate quantification could even be formulated.

See also: [Abstract objects](#); [Logical and mathematical terms, glossary of Universals](#)

LINDA WETZEL

References and further reading

- Block, N.** (ed.) (1980) *Readings in Philosophy of Psychology*, Cambridge, MA: Harvard University Press, vol.1. (Includes numerous important articles in philosophy of mind pertinent to the debate between token materialists and type materialists.)
- Bromberger, S.** (1992) 'Types and Tokens in Linguistics', in *On What We Know We Don't Know*, Chicago, IL: University of Chicago Press. (A fine attempt to explain the relationship between types and tokens, in order to explain how linguistics is an empirical science although about abstract types.)
- Chomsky, N.** (1957) *Syntactic Structures*, The Hague: Mouton. (The classic attack on structuralism which started the Chomskian revolution in linguistics.)
- Goodman, N.** (1951) *The Structure of Appearance*, Dordrecht: Reidel, 3rd edn, 1977. (Although not a book about types, Goodman's attempt to do without them produces many instructive passages.)
- Goodman, N. and Quine, W.V.** (1947) 'Steps Toward a Constructive Nominalism', *Journal of Symbolic Logic* 12: 105-22; repr. in *Problems and Projects*, Indianapolis, IN: Bobbs-Merrill, 1972. (A nominalist attempt at proof theory without expression types.)
- Hale, R.J.** (1987) *Abstract Objects*, Oxford and New York: Blackwell. (An excellent discussion about abstract objects generally, including types, and our reasons for being justified in concluding they exist.)
- Hutton, C.** (1990) *Abstraction and Instance: The Type-Token Relation in Linguistic Theory*, Oxford: Pergamon. (By means of an impressive survey of the relevant literature, Hutton argues that linguistic theories rely uncritically on the type/token distinction (and on related notions such as that of *langue-parole* and competence-performance) and that all such accounts are problematic.)
- Katz, J.J.** (1981) *Language and Other Abstract Objects*, Totowa, NJ: Rowman & Littlefield. (A robust defence of languages and sentences as abstract objects.)
- Peirce, C.S.** (1931-58) *Collected Papers of Charles Sanders Peirce*, ed. C. Hartshorne and P. Weiss, Cambridge, MA: Harvard University Press, vol. 4, para. 537; vol. 8, para. 334; vol. 2, para. 243-6. (The eight-volume collection of Peirce's papers.)
- Quine, W.V.** (1987) *Quiddities: An Intermittently Philosophical Dictionary*, Cambridge, MA: Harvard University Press, 216-19. (This is Quine distilled down to eighty-three alphabetically arranged entries, including a two-page entry on 'Type versus Token'.)
- Wetzel, L.** (1993) 'What Are Occurrences of Expressions?', *Journal of Philosophical Logic* 22: 215-20. (Provides an answer to the title question.)
- Wollheim, R.** (1968) *Art and Its Objects*, New York: Harper & Row. (Argues that many works of art cannot be physical objects/events or classes thereof, but are types that none the less are 'present in' their tokens.)
- Wolterstorff, N.** (1975) 'Toward an Ontology of Art Works', *Noûs* 9: 115-42. (Wolterstorff argues that many works of art are types, or 'kinds', and proposes a suitable analysis of the notion of a kind.)

Use/mention distinction and quotation

Speakers ‘use’ the expressions they utter and ‘mention’ the individuals they talk about. Connected with the roles of used expressions and mentioned individuals is a way of uniting them and a characteristic mistake involving them. Usually the expression used in an utterance will not be the same as the individual mentioned, but the two can be made to converge. The means is quotation. Quotation is a special usage in which an expression is used to mention itself. A failure to distinguish between the roles of used expressions and mentioned individuals can lead to mistakes. Such mistakes are called use/mention confusions. In themselves use/mention confusions are a minor linguistic faux pas, but under unfavourable conditions, they have the potential to cause greater problems.

1 Use and mention

Though philosophers since Plato have distinguished between language and the world it is used to describe, the terms ‘use’ and ‘mention’ were first applied to these categories by the philosopher and logician Willard Quine in 1940. Quine’s choice of the terms was intended to encourage associations with their ordinary meanings. To use an expression linguistically is to use it as a tool; to employ it towards a linguistic end. The ends are performances of speech acts, such as making claims, asking questions and giving responses. In contrast, to mention an individual is to say something about it; to make it the topic of conversation between speaker and listener. An individual mentioned in a speech act is one whose characteristics help determine how the act is to be evaluated: whether a claim is true, a question pertinent or a response correct depends on the individuals mentioned in the act and what is said about them.

In every communicative setting, some expression is used and, in most, some individual is mentioned. For example, a speaker uttering (1) below succeeds in making a claim about Tucson, the capital city of the state of Arizona:

(1) Tucson is dry.

By using the words ‘Tucson’, ‘is’ and ‘dry’ in the appropriate order, the city is mentioned. To say that Tucson is mentioned is to say that it is Tucson’s weather, rather than that of Rangoon or Vienna, which is relevant to the truth of the claim. Greetings, such as ‘Good Morning’ or ‘Hello’, are examples of speech acts in which no individual is mentioned; their success depends only on the circumstances and the speaker’s intentions.

While thinking about the categories of expressions used and individuals mentioned, one should bear in mind that the distinction divides the roles individuals can play rather than the kinds of individuals that can play these roles. That the distinction does not separate individuals by kinds is shown by the fact that linguistic expressions can play both parts: expressions can be mentioned as well as used. By uttering (2), for example, a speaker manages to mention the name ‘Tucson’ and, in fact, to say something true about it:

(2) The name of the capital city of Arizona contains six letters.

(1) should be compared with (2). The former uses the name of the city to mention the city, while the latter uses a description of the name to mention the name itself.

A detailed understanding of use and mention requires that one draw a distinction between the ‘type’ of an expression and ‘tokens’ of that type (see [Type/token distinction](#)). Expression tokens are concrete objects, individual marks or sounds made of raw materials: ink, perhaps; or sound waves. Types are abstract individuals and, as such, not made of anything. There is only one word ‘Tucson’ in the English language (so ‘Tucson’ is a type), of which there are several tokens in this entry.

The type/token distinction brings with it possibilities for what is used and mentioned. Both tokens and types can be used. To use a token, one has to utter it with the intention of saying something. Using a type requires the same intention, but also requires that the token have the right shape or sound. Tokens and types can also be mentioned.

(2) mentions the type and (3) the token used in (1):

(3) The first word token in the first example sentence of the entry on ‘Use/mention distinction and quotation’ in the *Routledge Encyclopedia of Philosophy* has six letters.

2 Standard uses of quotation and quasi-quotation

Expressions of each grammatical category have what may be called standard uses. The standard use of noun phrases is to mention: ‘Tucson’, ‘The name of the capital city of Arizona’ and ‘The first word token...*Philosophy*’ from (3) above are all used in standard fashion. Expressions of most other categories have standard uses which do not involve mentioning. Articles and quantifiers (‘every’, ‘some’, ‘all’, ‘most’) do not themselves mention, but rather help determine the range of individuals that are mentioned. Conjunctions govern the manner in which the meanings of compound expressions depend on the meanings of their constituents. According to some theories, adjectives, adverbs and verbs mention relations, properties or events. According to other theories, they do not mention at all.

Along with their standard uses, expressions also have nonstandard uses. Chief among these is their use in quotation. Quotation is a way of using an expression to mention itself. It is the most general and efficient way of talking about expressions. Instead of the long-winded description in (2), a speaker wanting to mention ‘Tucson’ can simply quote it:

(4) ‘Tucson’ has six letters.

From the notational point of view, quotation is trivial. Its basic rule is simple: to quote an expression, enclose a token of it in quotation marks - the purpose of the marks being to indicate the nature of the usage. Or it can be simpler yet; one can leave out the marks when it is clear from the context that one wants to talk about an expression rather than use it in the standard way. (Passages which are indented or examples marked by numbers - as in this entry - are quoted using this method.) Quotation is generally used to mention expression types, unless there is an explicit qualification to the contrary.

Quotation is a leveller. No matter what they may standardly be used to mention, when quoted, all nouns mention expressions. It is instructive to compare (1) and (4) on this count to see how quotation can bring about a change in subject matter: (1) is about a large city, while (4), which differs only in the addition of a set of quotation marks, is about a medium-sized word. The effect of quotation on other categories of expressions is even more significant. It reduces the whole variety of standard uses to one. No matter what their standard use may be, when quoted, expressions of all categories become mentioning expressions. This effect can be seen in the example below:

(5) ‘A’ is an article and ‘and’ is a conjunction.

In this sentence, the first token of the word ‘a’ and the second token of the word ‘and’ are used to mention, while the second token of the word ‘a’, the first of ‘and’ and the only one of ‘an’ operate in standard fashion.

Related to regular quotation is a notation invented by Quine (1940), called ‘quasi-quotation’. Quine, who has done more to encourage the correct use of quotation marks than any other writer, created quasi-quotation to help in capturing linguistic generalizations. Quasi-quotation treats an expression in two ways. Towards one part of the expression (‘Čamp;’ in the example below) it behaves as regular quotation. The other part or parts (‘ α ’ and ‘ β ’ in the example below) it treats as indeterminate elements that can mention any expression within a specific category.

The marks of quasi-quotation are called corner quotes (‘ $\lceil$ ’, on the left and ‘ $\rceil$ ’, on the right). A full quasi-quotation consisting of the expression enclosed by two corners is itself an indeterminate that can mention any expression consisting of the quoted part and any expression mentioned by the indeterminate part. To take an example from logic, let ‘ α ’ and ‘ β ’ be indeterminates mentioning sentences and ‘Čamp;’ be the conjunction symbol. A rule giving the extensional semantics for conjunction can be stated using quasi-quotation:

(6) $\lceil \alpha \ \& \ \beta \rceil$ is true iff α is true and β is true,

where ‘ $\lceil \alpha \ \& \ \beta \rceil$ ’ mentions an expression that is a conjunction of the sentences mentioned by ‘ α ’ and ‘ β ’. If α is ‘Cicero died’ and β is ‘Caesar spoke’, then $\lceil \alpha \ \& \ \beta \rceil$ is the result of writing α followed by ‘Čamp;’ followed by β , namely ‘Cicero died and Caesar spoke’.

3 Use/mention confusions

The distinction between use and mention can turn slippery. When it does and one loses hold of it, use/mention

confusions can occur. The typical result of such a confusion is that a speaker ends up making a claim about nonlinguistic individuals that is appropriate only to expressions or making a claim about expressions that is only appropriate to nonlinguistic individuals. Where the distinction between use and mention turns most slippery is on the subject of abstract individuals. Numbers and expressions used to mention them can be especially hard to keep apart. The category of numbers includes whole numbers such as 1 and 3, ratios such as $\frac{1}{2}$ and $\frac{4}{3}$ and irrational numbers such as π . The expressions that mention them include the numerals '1' and '3', the fractions ' $\frac{1}{2}$ ' and ' $\frac{4}{3}$ ', and the Greek letter ' π '. Classroom presentations of ratios and fractions commonly involve use/mention confusions. Students are often told something like the following:

Fractions consist of one number divided by another. The number to be divided, called the 'numerator', is placed on the top, while the number to divide by, called the 'denominator', is placed on the bottom, with a line between the two. So, 2 is the numerator and 3 the denominator of $\frac{2}{3}$.

However, the clever student may wonder how this could be. Since $\frac{2}{3}$ is the same quotient as $\frac{4}{6}$, the above remarks suggest that they ought to be the same fraction. Therefore 3 ought to be the denominator of both (on the grounds that the denominators of identicals are identical), but it is not. This puzzle exposes the use/mention confusion. The teacher has confused fractions with ratios. ' $\frac{2}{3}$ ' is a fraction; $\frac{2}{3}$ a ratio. Only fractions have numerators and denominators, while it is the ratios mentioned by fractions that may be identical. The numeral '3' is the denominator of the fraction ' $\frac{2}{3}$ ', but it is the ratios $\frac{2}{3}$ and $\frac{4}{6}$ which are identical. Because ' $\frac{2}{3}$ ' is not the same fraction as ' $\frac{4}{6}$ ', they need not, and indeed do not, have the same denominator.

The need to achieve clarity and avoid confusions such as the one above is the main reason why it is important to maintain a clear separation between use and mention. As far as errors go, the mix-up with fractions and ratios is probably mid-level. Some use/mention confusions are of less consequence. Others pose greater problems (see [Modal operators §1](#)). Where subtlety and precision are at a premium, as they are in logic and linguistics, use/mention confusions can severely impair the intelligibility of a discussion.

See also: [De re/de dicto](#); [Logical and mathematical terms, glossary of](#)

COREY WASHINGTON

References and further reading

- Davidson, D.** (1984) 'Quotation', in *Inquiries into Truth and Interpretation*, Oxford: Oxford University Press, 1991, 79-92. (Davidson puts forward the view that quotation marks are demonstratives, an alternative to the view advanced in §2.)
- Frege, G.** (1892) 'Über Sinn und Bedeutung', *Zeitschrift für Philosophie und philosophische Kritik* 100: 25-50; trans. 'On Sense and Reference', in A.P. Martinich (ed.) *The Philosophy of Language*, New York: Oxford University Press, 3rd edn, 1996. (Frege's discussion in this essay contains what may be the first application of the use/mention distinction in the context of a philosophical argument. Frege also introduces the view, stated in §2, that quotation marks are context markers and that a quoted word is used to mention itself.)
- Garcia-Carpentera, M.** (1994) 'Ostensive Signs: Against the Identity Theory of Quotation', *Journal of Philosophy* 91: 253-64. (Endorses Davidson's theory and defends it against criticisms.)
- Geach, P.** (1957) *Mental Acts*, London: Routledge & Kegan Paul, 79-82. (Clearest statement of the view, first formulated by Tarski, that quotations can be seen as descriptions.)
- Quine, W.V.** (1940) *Mathematical Logic*, New York: Norton; repr. Cambridge, MA: Harvard University Press, 1951, 23-37. (A careful discussion of the use/mention distinction, quotation and quasi-quotation.)
- Russell, B.A.W.** (1919) 'Descriptions', in *Introduction to Mathematical Philosophy*, London: Allen & Unwin, 167-80; repr. in A.P. Martinich (ed.) *The Philosophy of Language*, New York: Oxford University Press, 1985, 213-19; 3rd edn, 1996. (Contains numerous examples of use/mention confusions.)
- Tarski, A.** (1933) *Pojęcie prawdy w językach nauk dedukcyjnych*, Warsaw; trans. J.H. Woodger (1956), 'On the Concept of Truth in Formalized Languages', in *Logic, Semantics, Metamathematics*, ed. J. Corcoran, Indianapolis, IN: Hackett Publishing Company, 2nd edn, 1983, 152-278. (The classic work on the view that quotations are names.)
- Washington, C.** (1992) 'The Identity Theory of Quotation', *Journal of Philosophy* 91: 582-605. (Gives a detailed

description and defence of Frege's view.)

Vagueness

It seems obvious that there are vague ways of speaking and vague ways of thinking - saying that the weather is hot, for example. Common sense also has it that there is vagueness in the external world (although this is not the usual view in philosophy). Intuitively, clouds, for example, do not have sharp spatiotemporal boundaries. But the thesis that vagueness is real has spawned a number of deeply perplexing paradoxes and problems. There is no general agreement among philosophers about how to understand vagueness.

1 Vagueness in language and the world

Many linguistic terms are vague. But of what exactly does their vagueness consist? Consider, for example, the term 'bald'. This term has borderline cases of application. There are people to whom the term, as it is ordinarily used, neither clearly applies nor clearly fails to apply. In this respect, the term 'bald' is different from the term 'square root', say. There are no borderline square roots, nor could there be. Let us say that a general term (a word or phrase which potentially applies to many things) is intensionally vague (that is, vague with respect to its meaning) only if its meaning permits possible borderline cases of application. Then 'bald' is intensionally vague.

Vagueness can also be found in singular terms (words or phrases that potentially apply to single things). Consider, for example, 'the friend of Amy'. Suppose Amy has a love-hate relationship with Jane, without Amy clearly having any other friends. Then, it is indeterminate whether the singular term, 'the friend of Amy', designates Jane. So, 'the friend of Amy' has a possible borderline case of designation. And this indicates that it is intensionally vague.

It is plausible to suppose that intensional vagueness in general terms is more basic than intensional vagueness in singular terms, since uncontroversial examples of vague singular terms either contain vague general terms (non-relational or relational) or are the abstract singular counterparts to those terms. In the case of 'the friend of Amy', the term 'friend of' is intensionally vague.

The vagueness so far discussed does not require vagueness in the world. But worldly vagueness can be treated in a similar way. Suppose, as we normally do, that there really is a nonlinguistic property of being red, a property expressed by the predicate 'is red'. This predicate is vague and so is the property. Just as the former has borderline cases of application, so too the latter has borderline instances (objects which are neither clearly red nor clearly not red). Moreover, even if the actual world had not contained borderline red objects, still, intuitively, that would not have shown that the property, redness, is not vague. What the vagueness of properties seems to require is that there be possible borderline instances. Consider next concrete objects, for example, Mount Everest. Some molecules are definitely inside Everest and some are definitely outside. But intuitively, some have a borderline status: there is no determinate fact of the matter about whether they are inside or outside. Everest, then, has borderline spatiotemporal parts. In this way, it is like a cloud. So Everest is a vague concrete object.

There is another way in which, according to some philosophers, objects in the world can be vague. If, for a given object, o , there is an object, o' , such that it is indefinite whether o is identical with o' , then o , by virtue of its entering into an indefinite identity relation, is a vague object. Here is one possible example (due to D. Parfit, cited in [Broome 1984](#)). There is a club which has a clubhouse, a membership list and a set of rules. This club is never formally disbanded, but through time its members meet less and less frequently and the clubhouse becomes run down. There are no meetings for several years. Twelve years later, however, a few of the original members get together with some new people and start to meet once more in the same building (now redecorated). The club they belong to at this later date has the same name as the earlier one.

It has been held that the claim that the first club is identical with the second one is indefinite. Moreover, this indefiniteness, it has been suggested, is not an epistemic matter, since there is no further information which would settle the issue. On this view, each club has a vague identity and is thereby a vague object.

There is at least one respect in which the proposed characterizations of vagueness are incomplete. Consider the following sequence of conditionals, each of which includes the vague term 'tall':

(1a) If a man whose height is 7 feet is tall, then a man whose height is 6 feet 11 and $\frac{99}{100}$ inches is tall.

- (1b) If a man whose height is 6 feet 11 and $\frac{99}{100}$ inches is tall, then a man whose height is 6 feet 11 and $\frac{98}{100}$ inches is tall.
- ⋮
- (1n) If a man whose height is 5 feet and $\frac{1}{100}$ inches is tall, then a man whose height is 5 feet is tall.

Intuitively, there is no sharp dividing line between the true conditionals in this sequence and those that have some other value. This fact is not captured in the earlier account of intensional vagueness. But it seems to be part and parcel of our ordinary conception of the vagueness of terms such as 'tall' or 'bald'. Their vagueness is robust or resilient. And a corresponding point can be made about vague concrete objects. For example, it does not seem to be true that there is a sharp dividing line between the molecules that are inside Everest and those that have a borderline status.

2 Paradoxes and problems

The oldest puzzle of vagueness, which allegedly derives from Eubulides, is the paradox of the bald man. It goes as follows:

- (2) A man with no hairs on his head is bald.
 (3) For any number, n , if a man with n hairs on his head is bald then a man with $n + 1$ hairs on his head is bald.
 (4) Therefore, a man with ten thousand hairs on his head is bald.

The conclusion is derived from the premises via ten thousand applications of the classical logical rules of *modus ponens* and universal instantiation, so this argument is valid. Now, premise (2) is certainly true and the conclusion, (4), is certainly false, so it is inferred that premise (3) is false. Therefore, there is an n such that a man with n hairs on his head is bald and a man with $n + 1$ is not. Therefore, the term 'bald' is precise, contrary to appearances. And what is true in this one case is true by parallel reasoning for any general term which is ordinarily classified as vague.

It is worth observing that the paradox of the bald man does not rely essentially on the use of a universal generalization. The same result can be generated by means of a sequence of (ten thousand) conditionals such as in (1a)-(1n), with (2) as a starting point. Since the conclusion, (4), is false, and (2) is true, then at least one of the conditionals must be false. So, given that (2) is true and that each conditional is either true or false, there must be an adjacent pair of conditionals such that the first is true and the second false. This runs directly counter to the idea that 'bald' is vague.

If the paradox of the bald man really does demonstrate that the term 'bald' is not vague, contrary to what we all ordinarily believe, then 'bald' cannot express a vague property. So, the property of being bald cannot itself be vague. Since the paradox of the bald man can be restated so as to apply to any vague general term, one overall conclusion which has been reached is that no general terms or properties are vague. This conclusion, of course, not only threatens the thesis that there are vague abstract objects but also challenges the idea that there is any such thing as vagueness in language at all, however it is understood.

There are also sorites arguments which attack the concrete objects of common sense, on the assumption that these objects have vague boundaries. These arguments, like the paradox of the bald man, can be stated without the use of a universal generalization.

There are other problems for the view that individual concrete objects are vague. One of these is directed against vague objects, conceived of as objects having indeterminate identities (Evans 1978). Suppose that ' a ' and ' b ' are (precise) singular terms for vague objects and that ' $a = b$ ' is indefinite in truth-value. Then, if we let ' ∇ ' symbolize 'indefinitely', the following is true:

- (5) $\nabla(a = b)$

(5) ascribes to b the property of being indefinitely identical with a . Now surely we have

- (6) $\sim\nabla(a = a)$

and hence that a lacks the property of being indefinitely identical with a . By the principle that if object o and object o' differ in a property, then they are not identical (the contrapositive of Leibniz's Law; see **Identity**), it follows that a is not identical with b . Since, on the standard understanding of 'indefinitely', it cannot be the case both that it is indefinite whether a is identical with b and that a is not identical with b , there cannot be vague identities.

The deepest problems of vagueness, in my view, are the sorites paradoxes. And these arise whether or not there is vagueness in the nonlinguistic world.

3 Theories

It has been suggested that a proper understanding of vagueness requires the admission that truth and set membership come in degrees, rather than being 'all or nothing' (see **Zadeh 1965**, **Goguen 1969**). On this approach, real numbers in the interval from 0 to 1 are typically taken to be truth-values, with 1 being fully fledged truth and 0 being fully fledged falsity. The same numbers are assigned to the degrees to which objects belong to sets. So, for example, if Herbert is clearly bald, then the sentence

(7) Herbert is bald

is assigned the value 1, and Herbert is taken to belong to the set of bald men to degree 1. But if Herbert is a borderline case then (7) is assigned some number less than 1, say, 0.6, and the degree of Herbert's membership in the set of bald men is now taken to be 0.6 too. In general, a singular sentence ' Fa ' is treated as having the truth-value n , where $0 \leq n \leq 1$, if and only if the referent of ' a ' belongs to the set of F s to degree n (see **Fuzzy logic**).

A consequence of this approach is that vague predicates do not sharply divide the world into those things to which they apply and those things to which they do not. Rather, vague predicates apply to objects to varying degrees. As a man loses hair he becomes more and more bald; so the predicate 'is bald' applies to him more and more and the assertion that he is bald increases gradually in its degree of truth.

One serious objection to this view is that it really replaces vagueness with the most refined and incredible precision. Set membership, as viewed by the degrees-of-truth theorist, comes in precise degrees, as does predicate application and truth. The result is a commitment to precise dividing lines that is not only unbelievable but also thoroughly contrary to how we think of everyday vagueness (as noted in §1).

It is often supposed by philosophers that vagueness resides in language. One popular view here is that vagueness itself is a matter of semantic indecision (for example, **Lewis 1986**). On this view, all external objects, properties and relations are precise. Each vague linguistic term has a meaning which can be made precise in a whole range of different permissible ways. For example, 'over 70 years of age' is one acceptable way of making precise the meaning of the term 'old'. But other equally good ways are 'over 72' and 'over 74'. The rules that govern the use of the term 'old' are not specific enough for us to be able to choose non-arbitrarily some one precise term as capturing its meaning. When the term 'old' was first introduced, no decision was made to link it exclusively with one particular precise property. Instead it was applied widely in a range of cases, thereby acquiring a vague meaning. And what is true here for 'old' is true *mutatis mutandis* for all other vague terms.

Now if it is indeed true that there are no vague properties expressed by vague predicates, but instead ranges of precise properties, then, under some ways of making precise a vague predicate, it will definitely apply to a given object (in virtue of the object having the appropriate precise property) and, under other ways, it will not. Suppose that Alfred is 67. If 'old' is sharpened to mean 'over 70' then Alfred is not old: he lacks the precise property of being over 70. But if 'old' is sharpened to mean 'over 65', then Alfred is old - this time he has the right precise property. Is it really true that Alfred is old, then? If these two ways of sharpening the term 'old' are equally acceptable - if they both express precise properties falling in the range associated with 'old' - then the natural answer is that it is neither true nor false that Alfred is old. He is a borderline case. So, there really is no definite age at which Alfred becomes old: the boundary between being old and not being old is not sharp. By contrast, if Alfred is 97 years of age, then, under any acceptable way of making 'old' precise, he will count as old. So here it is unquestionably true that Alfred is old. And if he is 21, then, no matter which acceptable precisification we choose for 'old', he will not count as old. The claim that Alfred is old is false.

The thesis that vagueness is semantic indecision thus leads straightforwardly to the following claims. Vague sentences have three possible truth-values: true, false and ‘indefinite’ (neither true nor false). A vague sentence is to be counted as true if it comes out true under all acceptable ways of making precise its component vague terms; as false if it comes out false in every such case; and as neither true nor false if it comes out true under some ways and false under others. These claims entail that the law of excluded middle (LEM) remains true for vague sentences. So, the semantic indecision theory of vagueness provides us with a classical conception of reality (all that there is in the world is precise), a safe haven for vagueness in the linguistic realm and a conservative attitude towards the retention of LEM.

There are grave difficulties, however. In particular, it is far from clear that the semantic indecision approach fares any better than the degrees-of-truth view with respect to sorites paradoxes and robust or resilient vagueness. Another objection concerns LEM. The objection is that

(8) Either Herbert is bald or Herbert is not bald,

to take one instance of LEM, should not be counted as true if Herbert is a borderline bald man. For if (8) is true then either Herbert is bald or he is not bald. If this is the case then the question, ‘Well, which is he then?’, must surely have an answer; that is, if (8) is true then precisely one of the disjuncts in (8) must be true. So ‘Herbert is bald’ must be either true or false. This runs contrary to the assumption that Herbert is a borderline bald man. So, (8) is not true.

One common feature of both the standard degrees-of-truth approach and the semantic indecision theory is that they employ precise metalanguages in stating the truth-conditions for vague object language sentences (a metalanguage is used to talk about the object language). It is this feature which, in my view, dooms them to failure. A central aspect of the resilient vagueness of ordinary terms is that, in sorites sequences, there simply is no determinate fact of the matter about the transition from true to some other value. Any attempt to state truth-conditions for vague discourse in precise language will inevitably fall foul of this fact.

Another possible approach, then, is to start out using a vague metalanguage, which mirrors in its vagueness the object language. One proposal along these lines is based on the three truth-values, true, false and indefinite. Just as sentences in the object language can be indefinite, so too can sentences in the metalanguage (and indeed in all the higher metalanguages). If this approach is to avoid the sorites paradoxes, it is essential that it not be true that every object language sentence is true or false or indefinite. For this would create sharp dividing lines. But neither can it be false that every such sentence is true or false or indefinite. For this would require further truth-values. Instead, it has been argued that the above generalization about object language sentences must be indefinite. This permits us to hold that the claim that there is a last true sentence, followed by a first indefinite one, in a sorites sequence is indefinite. And if this is the case then sorites arguments cannot be sound (see [Tye 1990](#)).

There are many philosophers who would reject all three of the alternative proposals sketched so far. The fourth and final approach I shall mention is the epistemic view ([Sorensen 1988](#), [Williamson 1994](#)). According to this position, vagueness is a kind of ignorance. The nonlinguistic world is precise. Standard logic holds even for vague discourse. Every molecule is either inside or outside Everest, for example. We just do not know where the boundaries lie. Likewise, there is always a single hair, the addition of which would turn a bald man into a man who is not bald, even though we cannot say which hair this is. Our sensory and conceptual mechanisms are simply not equipped to make the necessary fine-grained discriminations. Sorites arguments, then, rest upon a false premise.

This view is pleasingly straightforward. Unfortunately, it seems counterintuitive. It denies outright the existence of robust or resilient vagueness. Moreover, it seems to misconceive borderline cases. Our ordinary concept of a borderline case is the concept of a case that is neither one thing nor the other. ‘Definitely’ here does not seem to mean ‘known’ or ‘knowably’.

So, whichever way we turn, we quickly become enmeshed in difficulties. Of all the philosophical mysteries, vagueness is one of the most perplexing.

See also: [Many-valued logics, philosophical issues in](#); [Logical and mathematical terms, glossary of](#)

MICHAEL TYE

References and further reading

- Broome, J.** (1984) 'Indefiniteness in Identity', *Analysis* 44: 6-12. (Discussion of whether objects can have vague identities.)
- Dummett, M.** (1975) 'Wang's Paradox', *Synthese* 25: 301-24. (Vague observational predicates as logically incoherent.)
- Evans, G.** (1978) 'Can There Be Vague Objects?', *Analysis* 38: 208. (An argument against the existence of vague objects.)
- Fine, K.** (1975) 'Vagueness, Truth, and Logic', *Synthese* 25: 265-300. (Classic early presentation of vagueness as semantic indecision.)
- Forbes, G.** (1985) *The Metaphysics of Modality*, Oxford: Clarendon Press. (Includes a version of the degrees-of-truth approach.)
- Goguen, J.** (1969) 'The Logic of Inexact Concepts', *Synthese* 19: 325-73. (An early statement of the degrees-of-truth theory.)
- Heller, M.** (1990) *The Ontology of Physical Objects: Four-Dimensional Hunks of Matter*, Cambridge: Cambridge University Press. (Sorites arguments deployed in defence of the view that the world is precise.)
- Horgan, T. (ed.)** (1994a) *Spindel Conference on Vagueness*, supplement to *Southern Journal of Philosophy* 33. (Good collection of new papers on vagueness.)
- Horgan, T. (ed.)** (1994b) 'Robust Vagueness and the Forced March Sorites Paradox', *Philosophical Perspectives* 8: 159-88. (A new form of sorites reasoning and its consequences.)
- Inwagen, P. van** (1988) 'How To Reason About Vague Objects', *Philosophical Topics* 16: 255-84. (A non-classical logic for reasoning about vague objects.)
- Lewis, D.K.** (1986) *The Plurality of Worlds*, Oxford: Blackwell. (Simple statement of the semantic indecision theory.)
- Lewis, D.K.** (1993) 'Many, But Almost One', in K. Campbell, J. Bacon and L. Reinhardt (eds) *Ontology, Causality, and Mind: Essays on the Philosophy of D.M. Armstrong*, Cambridge: Cambridge University Press. (Arguments against the existence of vague objects.)
- Putnam, H.** (1983) 'Vagueness and Alternative Logic', *Erkenntnis* 19: 297-314. (A discussion of vagueness and non-classical logic.)
- Sainsbury, M.** (1991) 'Is There Higher-Order Vagueness?', *Philosophical Quarterly* 41: 167-82. (Discussion of whether there can be borderline borderline cases.)
- Sanford, D.** (1975) 'Borderline Logic', *American Philosophical Quarterly* 12: 29-39. (Formal development of the degrees-of-truth approach.)
- Sorensen, R.** (1988) *Blindspots*, Oxford: Clarendon Press. (Vagueness as ignorance.)
- Thomason, R.** (1982) 'Identity and Vagueness', *Philosophical Studies* 42: 329-32. (Discussion of Evans' argument from identity against the existence of vague objects.)
- Tye, M.** (1990) 'Vague Objects', *Mind* 99: 535-57. (Vagueness in both the object language and the metalanguage; defence of vagueness in the world.)
- Unger, P.** (1979a) 'There are No Ordinary Things', *Synthese* 41: 117-54. (Sorites arguments against the existence of ordinary things.)
- Unger, P.** (1979b) 'Why There are No People', *Midwest Studies in Philosophy* 4: 177-222. (Sorites arguments against the existence of people.)
- Wheeler, S.** (1979) 'On That Which is Not', *Synthese* 41: 155-73. (Sorites arguments against the existence of ordinary things.)
- Williamson, T.** (1994) *Vagueness*, London: Routledge. (Clear defence of the view that vagueness is ignorance.)
- Williamson, T.** (1975) 'On the Coherence of Vague Predicates', *Synthese* 30: 325-63. (Broad examination of the nature of vagueness.)
- Zadeh, L.** (1965) 'Fuzzy Sets', *Information and Control* 8: 338-53. (Early presentation of the degrees-of-truth approach.)

