WILFRID HODGES

ELEMENTARY PREDICATE LOGIC

INTRODUCTION

Elementary (first-order) predicate logic is a child of many parents. At least three different groups of thinkers played their part in its conception, with three quite distinct motives. Maybe the mixture gave it hybrid strength. But whatever the reason, first-order logic is both the simplest, the most powerful and the most applicable branch of modern logic.

The first group who can claim paternity are the *Traditional Logicians*. For these scholars the central aim of logic was to schematise valid arguments. For present purposes an argument consists of a string of sentences called *premises*, followed by the word '*Therefore*', followed by a single sentence called the *conclusion*. An argument is called *valid* when its premises *entail* its conclusion, in other words, if the premises can't be true without the conclusion also being true.

A typical valid argument schema might be:

1. a is more X than b. b is more X than c.
   Therefore a is more X than c.

This becomes a valid argument whenever we substitute names for a, b, c respectively and an adjective for X; as for example

2. Oslo is more clean than Ydstebøhavn. Ydstebøhavn is more clean than Trondheim. *Therefore* Oslo is more clean than Trondheim.

Arguments like (2) which result from such substitutions are called *instances* of the schema (1). Traditional logicians collected valid argument schemas such as (1). This activity used to be known as *formal logic* on the grounds that it was concerned with the forms of arguments. (Today we more often speak of formal versus informal logic, just as formal versus informal semantics, meaning mathematically precise versus mathematically imprecise.)

The ancients and the medievals had concerned themselves with small numbers of argument schemas gathered more or less *ad hoc*. Aristotle's syllogisms give twenty-four schemas, of which Aristotle himself mentions nineteen. The watershed between classical and modern logic lies in 1847, when George Boole (1815–1864) published a calculus which yielded infinitely many valid argument schemas of arbitrarily high complexity (Boole [1847; 1854]). Today we know Boole's calculus as *propositional logic*. Other early researchers who belong among the Traditionals are Augustus De Morgan (1806–1871) and C. S. Peirce (1839–1914). Their writings are lively with examples of people i being enemies to people j at time k, and other people overdrawing their bank accounts.

The second group of originators were the *Proof Theorists*. Among these should be included Gottlob Frege (1848–1925), Giuseppe Peano (1858–1932), David Hilbert (1862–1943), Bertrand Russell (1872–1970), Jacques Herbrand (1908–1931) and Gerhard Gentzen (1909–1945). Their aim was to systematise mathematical reasoning so that all assumptions were made explicit and all steps rigorous. For Frege this was a matter of integrity and mental hygiene. For Hilbert the aim was to make mathematical reasoning itself the object of mathematical study, partly in order to justify infinitary mathematics but partly also as a new method of mathematical research. This group devised both the notation and the proof theory of first-order logic. The earliest calculus adequate for first-order logic was the system which Frege published in his *Begriffschrift* [1879]. This was also the first work to discuss quantifiers.

With a slight anachronism I call the third group the *Model Theorists*. Their aim was to study mathematical structures from the point of view of the laws which these structures obey. The group includes Ernst Schröder (1841–1902), Leopold Löwenheim (1878–1957), Thoralf Skolem (1887–1963), C. H. Langford (1895?–1964), Kurt Gödel (1906–1978) and Alfred Tarski (1901–1983). The notion of a first-order property is already clear in Schröder's work [1895], though the earliest use I could find of the term 'first-order' in the modern sense is in Langford [1927]. (Langford quotes the slightly different use of the term *Principia Mathematica*, Whitehead and Russell [1910].)

Our present understanding of what first-order logic is about was painstakingly built up by this group of workers during the years 1915 to 1935. The progress was conceptual as much as technical; a historian of logic feels his fingers tingle as he watches it. Increasing precision was an important part of it. But it is worth reflecting that by 1935 a logician could safely say 'The formal sentence S is true in the structure A' *and mean it.* Frege [1906] had found such language morally reprehensible (cf. Section 12 below). Skolem [1922] talked of formal axioms 'holding in a domain', but he felt obliged to add that this was 'only a manner of speaking, which can lead only to purely formal propositions—perhaps made up of very beautiful *words*...'. (On taking truth literally, see above all Kurt Gödel's letters to Hao Wang, [1974, p. 8 ff] and the analysis by Solomon Feferman [1984]. R. L. Vaught's historical paper [1974] is also valuable.)

Other groups with other aims have arisen more recently and found first-order logic helpful for their purposes. Let me mention two.

One group (if we can lump together such a vast army of workers) are the computer scientists. There is wide agreement that trainee computer scientists need to study logic, and a range of textbooks have come onto the market aimed specifically at them. (To mention just two, Reeves and

Clarke [1990] is an introductory text and Gallier [1986] is more advanced.) But this is mainly for training; first-order logic itself is not the logic of choice for many computer science applications. The artificial intelligence community consume logics on a grand scale, but they tend to prefer logics which are modal or intensional. By and large, specification languages need to be able to define functions, and this forces them to incorporate some higher-order features. Very often the structures which concern a computer scientist are finite, and (as Yuri Gurevich [1984] argued) first-order logic seems not to be the best logic for classifying finite structures.

Computer science has raised several questions which cast fresh light on first-order logic. For example, how does one search for a proof? The question itself is not new—philosophers from Aristotle to Leibniz considered it. What is completely new is the mathematical analysis of systematic searches through all possible proofs in a formal calculus. Searches of this kind arise naturally in automated theorem proving. Robert Kowalski [1979] proposed that one could read some first-order sentences as instructions to search for a proof; the standard interpretation of the programming language PROLOG rests on his idea. Another question is the cost of a formal proof, in terms of the number of assumptions which are needed and the number of times each assumption is used; this line of enquiry has led to fragments of first-order logic in which one has some control over the cost (see for example Jean-Yves Girard [1987; 1995] on linear logic and Došen and Schroeder-Heister [1993] on substructural logics in general).

Last but in no way least come the linguists. After Chomsky had revolutionised the study of syntax of natural languages in the 1950s and 60s, many linguists shifted the spotlight from grammar to meaning. It was natural to presume that the meaning of a sentence in a natural language is built up from the meanings of its component words in a way which reflects the grammatical structure of the sentence. The problem then is to describe the structure of meanings. One can see the beginnings of this enterprise in Bertrand Russell's theory of propositions and the 'logical forms' beloved of English philosophers earlier in this century; but the aims of these early investigations were not often clearly articulated. Round about 1970 the *generative semanticists* (we may cite G. Lakoff and J. D. McCawley) began to use apparatus from first-order logic in their analyses of natural language sentences; some of their analyses looked very much like the formulas which an up-to-date Traditional Logician might write down in the course of knocking arguments into tractable forms. Then Richard Montague [1974] opened a fruitful line of research by using tools from logic to give extremely precise analyses of both the grammar and semantics of some fragments of English. (Cf. Dowty *et al.* [1981] for an introduction to Montague grammar.) I should add that many researchers on natural language semantics, from Montague onwards, have found that they needed logical devices which go far beyond first-order logic. More recently some of the apparatus of first-order proof theory has turned up unexpectedly in the analysis of grammar; see for example Morrill [1994] and Kempson [1995].

Logicians like to debate over coffee when 'real' first-order logic first appeared in print. The earliest textbook account was in the *Grundzüge der theoretischen Logik* of Hilbert and Ackermann [1928], based on Hilbert's lectures of 1917–1922. Skolem's paper [1920] is undeniably about first-order logic. But Whitehead and Russell's *Principia Mathematica* [1910] belongs to an earlier era. It contains notation, axioms and theorems which we now regard as part of first-order logic, and for this reason it was quoted as a reference by Post, Langford, Herbrand and Gödel up to 1931, when it figured in the title of Gödel's famous paper on incompleteness, [Gödel, 1931b]. But the first-order part of *Prinicipia* is not distinguished from the rest; and more important, its authors had no notion of a precise syntax or the interpretation of formulas in structures.

I: Propositional Logic

1 TRUTH FUNCTORS

In propositional logic we use six artificial symbols $\neg, \wedge, \vee, \rightarrow, \leftrightarrow, \bot$, called *truth-functors*. These symbols all have agreed meanings. They can be used in English, or they can have an artificial language built around them.

Let me explain one of these symbols, $\wedge$, quite carefully. The remainder will then be easy.

We use $\wedge$ between sentences ϕ, ψ to form a new sentence

(1) $(\phi \wedge \psi)$.

The brackets are an essential part of the notation. Here and below, 'sentence' means 'indicative sentence'. If ϕ and ψ are sentences, then in any situation,

(2) $(\phi \wedge \psi)$ is true iff ϕ is true and ψ is true; otherwise it is false.

('Iff' means 'if and only if'.) This defines the meaning of $\wedge$.

Several points about this definition call for comment. First, we had to mention the situation, because a sentence can be true in one situation and not true in another. For example, the sentence may contain demonstrative pronouns or other indexicals that need to be given a reference, or words that need to be disambiguated. (The situation is not necessarily the 'context of utterance'—a sentence can be true in situations where it is never uttered.)

In propositional logic we assume that in every situation, each sentence under discussion is determinately either true or false and not both. This assumption is completely innocent. We can make it correct by adopting either or both of the following conventions. First, we can agree that although we intend to use the word 'true' as it is normally used, we shall take 'false' to mean simply 'not true'. And second, we can take it as understood that the term 'situation' covers only situations in which the relevant sentences are either true or false and not both. (We may also wish to put an embargo on nonsensical sentences, but this is not necessary.) There are of course several ways of being not true, but propositional logic doesn't distinguish between them.

Logicians always make one further assumption here: they assume that truth and falsehood—T and F for short—are objects. Then they say that the *truth-value* of a sentence is T if the sentence is true, and F otherwise. (Frege [1912]: '...in logic we have only two objects, in the first place: the two truth-values.') But I think in fact even the most scrupulous sceptic could follow the literature if he *defined* the truth-value of all true sentences to be his left big toe and that of false sentences to be his right. Many writers take truth to be the number 1, which they identify with the set $\{0\}$, and falsehood to be the number 0, which is identified with the empty set. Nobody is obliged to follow these choices, but technically they are very convenient. For example (2) says that if the truth-value of ϕ is x and the truth-value of ψ is y, then that of $(\phi \wedge \psi)$ is xy.

With this notation, the definition (2) of the meaning of $\wedge$ can be written in a self-explanatory chart:

(3)

ϕ	ψ	$(\phi \wedge \psi)$
T	T	T
T	F	F
F	T	F
F	F	F

The diagram (3) is called the *truth-table* of $\wedge$. Truth-tables were first introduced by C. S. Peirce in [1902].

Does (3) really define the meaning of $\wedge$? Couldn't there be two symbols $\wedge_1$ and $\wedge_2$ with different meanings, which both satisfied (3)?

The answer is that there certainly can be. For example, if $\wedge_1$ is any symbol whose meaning agrees with (3), then we can introduce another such symbol $\wedge_2$ by declaring that $(\phi \wedge_2 \psi)$ shall mean the same as the sentence

(4) $(\phi \wedge_1 \psi)$ and the number π is irrational.

(Wittgenstein [1910] said that $\wedge_1$ and $\wedge_2$ then mean the same! *Tractatus* 4.46ff, 4.465 in particular.) But this is the wrong way to read (3). Diagram (3) should be read as stating *what one has to check in order to determine that* $(\phi \wedge \psi)$ *is true.* One can verify that $(\phi \wedge \psi)$ is true without knowing that π is irrational, but not without verifying that ϕ and ψ are true. (See

Michael Dummett [1958/59; 1975] on the relation between meaning and truth-conditions.)

Some logicians have claimed that the sentence $(\phi \wedge \psi)$ means the same as the sentence

(5) ϕ and ψ.

Is this correct? Obviously the meanings are very close. But there are some apparent differences. For example, consider Mr Slippery who said in a court of law:

(6) I heard a shot and I saw the girl fall.

when the facts are that he saw the girl fall and *then* heard the shot. Under these circumstances

(7) (I heard a shot $\wedge$ I saw the girl fall)

was true, but Mr Slippery could still get himself locked up for perjury. One might maintain that (6) does mean the same as (7) and was equally true, but that the conventions of normal discourse would have led Mr Slippery to choose a different sentence from (6) if he had not wanted to mislead the jury. (See Grice [1975] for these conventions; Cohen [1971] discusses the connection with truth-tables.)

Assuming, then, that the truth-table (3) does adequately define the meaning of $\wedge$, we can define the meanings of the remaining truth-functors in the same way. For convenience I repeat the table for $\wedge$.

(8)

ϕ	ψ	$\neg\phi$	$\phi \wedge \psi$	$\phi \vee \psi$	$\phi \to \psi$	$\phi \leftrightarrow \psi$	$\bot$
T	T	F	T	T	T	T	F
T	F		F	T	F	F	
F	T	T	F	T	T	F	
F	F		F	F	T	T	

$\neg\phi$ is read 'Not ϕ' and called the *negation* of ϕ. $(\phi \wedge \psi)$ is read 'ϕ and ψ' and called the *conjunction* of ϕ and ψ, with *conjuncts* ϕ and ψ. $(\phi \vee \psi)$ is read 'ϕ or ψ' and called the *disjunction* of ϕ and ψ, with *disjuncts* ϕ and ψ. $(\phi \to \psi)$ is read 'If ϕ then ψ' or 'ϕ arrow ψ'; it is called a *material implication* with *antecedent* ϕ and *consequent* ψ. $(\phi \leftrightarrow \psi)$ is read 'ϕ if and only if ψ', and is called the *biconditional* of ϕ and ψ. The symbol $\bot$ is read as 'absurdity', and it forms a sentence by itself; this sentence is false in all situations.

There are some alternative notations in common use; for example

(9)
$-\phi$ or $\sim\phi$ for $\neg\phi$.
$(\phi \& \psi)$ for $(\phi \wedge \psi)$.
$(\phi \supset \psi)$ for $(\phi \to \psi)$.
$(\phi \equiv \psi)$ for $(\phi \leftrightarrow \psi)$.

Also the truth-functor symbols are often used for other purposes. For example the intutionists use the symbols $\neg, \wedge, \vee, \rightarrow, \leftrightarrow$ but not with the meanings given in (8); cf. van Dalen's chapter on Intutionistic Logic in a later volume. Some writers use the symbol $\rightarrow$ for other kinds of implication, or even as a shorthand for the English words 'If ... then'.

A remark on metavariables. The symbols 'ϕ' and 'ψ' are not themselves sentences and are not the names of particular sentences. They are used as above, for making statements about any and all sentences. Symbols used in this way are called *(sentence) metavariables.* They are part of the *metalanguage*, i.e. the language we use for talking about formulas. I follow the convention that when we talk about a formula, symbols which are not metavariables are used as names for themselves. So for example the expression in line (1) means the same as: the formula consisting of '(' followed by ϕ followed by '$\wedge$' followed by ψ followed by ')'. I use quotation marks only when clarity or style demand them. These conventions, which are normal in mathematical writing, cut down the clutter but put some obligation on reader and writer to watch for ambiguities and be sensible about them. Sometimes a more rigorous convention is needed. Quine's corners $\ulcorner\urcorner$ supply one; see Quine [1940, Section 6]. There are some more remarks about notation in Section 4 below.

2 PROPOSITIONAL ARGUMENTS

Besides the truth-functors, propositional logic uses a second kind of symbol, namely the *sentence letters*

(10) $p, q, r, \ldots, p_1, p_2, \ldots, .$

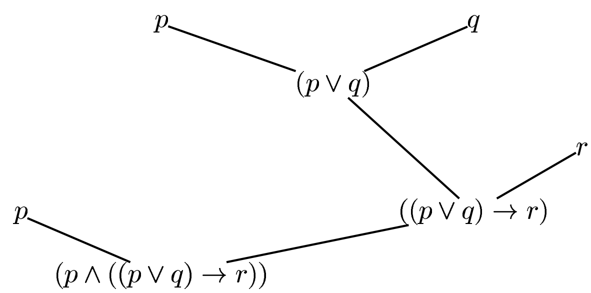
These letters have no fixed meaning. They serve to mark spaces where English sentences can be written. We can combine them with the truth-functors to produce expressions called *formulas*, which become sentences when the sentence letters are replaced by sentences.

For example, from the sentence letters p, q and r we can build up the formula

(11) $(p \wedge ((p \vee q) \rightarrow r))$

as follows:

(12)



We call (12) the *formation tree* of the formula (11). Sentence letters themselves are reckoned to be *atomic formulas*, while formulas which use truth-functors are called *compound formulas*. In a compound formula there is always a truth-functor which was added last in the formation tree; this occurrence of the truth-functor is called the *main connective* of the formula. In (11) the main connective is the occurrence of $\wedge$. The main connective of $\bot$ is reckoned to be $\bot$ itself.

Suppose ϕ is a formula. An *instance* of ϕ is a sentence which is got from ϕ by replacing each sentence letter in ϕ by an English sentence, in such a way that no sentence letter gets replaced by different sentences at different occurrences. (Henceforth, the symbols 'ϕ', 'ψ' are metavariables for formulas as well as sentences. The letters 'p', 'q' etc. are not metavariables; they are the actual symbols of propositional logic.)

Now if we know the truth-values of the inserted sentences in an instance of ϕ, then we can work out by table (8) what the truth-value of the whole instance must be. Taking (11) as an example, consider the following table:

(13)

	p	q	r	$(p$	$\wedge$	$((p$	$\vee$	$q)$	$\to$	$r))$
(*i*)	T	T	T	T	T	T	T	T	T	T
(*ii*)	T	T	F	T	F	T	T	T	F	F
(*iii*)	T	F	T	T	T	T	T	F	T	T
(*iv*)	T	F	F	T	F	T	T	F	F	F
(*v*)	F	T	T	F	F	F	T	T	T	T
(*vi*)	F	T	F	F	F	F	T	T	F	F
(*vii*)	F	F	T	F	F	F	F	F	T	T
(*viii*)	F	F	F	F	F	F	F	F	T	F
				1	7	2	5	3	6	4

The rows (i)–(viii) on the left list all the possible ways in which the sentences put for p and q can have truth-values. The columns on the right are computed in the order shown by the numbers at the bottom. (The numbers at left and bottom are not normally written—I put them in to help the explanation.) Columns 1, 2, 3, 4 just repeat the columns on the left. Column 5 shows the truth-value of $(p \vee q)$, and is calculated from columns 2 and 3 by means of table (8). Then column 6 is worked out from columns 5 and

4, using the truth-table for $(\phi \to \psi)$ in (8). Finally, column 7 comes from columns 1 and 6 by the table for $(\phi \wedge \psi)$. Column 7 is written under the main connective of (11) and shows the truth-value of the whole instance of (11) under each of the eight possibilities listed on the left.

Table (13) is called the *truth-table* of the formula (11). As we constructed it, we were working out truth-tables for all the formulas shown in the formation tree (12), starting at the top and working downwards.

We are now equipped to use propositional logic to prove the validity of an argument. Consider:

(14) That was a hornet, and soda only makes hornet and wasp stings worse. So you don't want to use soda.

This contains an argument along the following lines:

(15) (You were stung by a hornet ∧ ((you were stung by a hornet ∨ you were stung by a wasp) → soda will make the sting worse)). *Therefore* soda will make the sting worse.

We replace the component sentences by letters according to the scheme:

(16) p : You were stung by a hornet.
q : You were stung by a wasp.
r : Soda will make the sting worse.

The result is:

(17) $(p \wedge ((p \vee q) \to r))$. *Therefore* r.

Then we calculate truth-tables for both premise and conclusion of (17) at the same time. Only the main columns are shown below.

(18)

	p	q	r	$(p \wedge ((p \vee q) \to r))$.	*Therefore* r
(*i*)	T	T	T	T	T
(*ii*)	T	T	F	F	F
(*iii*)	T	F	T	T	T
(*iv*)	T	F	F	F	F
(*v*)	F	T	T	F	T
(*vi*)	F	T	F	F	F
(*vii*)	F	F	T	F	T
(*viii*)	F	F	F	F	F

Table (18) shows that if the premise of (15) is true then so is the conclusion. For if the premise is true, then the column under the premise shows that we are in row (i) or row (iii). In both of these rows, the last column in (18) shows that the conclusion is true. There is no row which has a T below $(p \wedge ((p \vee q) \to r))$ and an F below r. Hence, (15) is valid.

In the language of the traditional logician, these calculations showed that (17) is a valid argument schema. Every instance of (17) is a valid argument.

Note how the proof of the validity of an argument falls into two parts. The first is to translate the argument into the symbols of propositional logic. This involves no calculation, though a gauche translation can frustrate the second part. I say no more about this first part—the elementary textbooks give hundreds of examples [Kalish and Montague, 1964; Mates, 1965; Thomason, 1970; Hodges, 1977]. The second part of the proof is pure mechanical calculation using the truth-table definitions of the truth-functors. What remains to discuss below is the theory behind this mechanical part.

First and foremost, why does it work?

3 WHY TRUTH-TABLES WORK

If ϕ is any formula of propositional logic, then any assignment of truth-values to the sentence letters which occur in ϕ can be extended, by means of the truth-table definitions of the truth-functors, to give a truth-value to ϕ; this truth-value assigned to ϕ is uniquely determined and it can be computed mechanically.

This is the central thesis of propositional logic. In Section 2 I showed how the assignment to ϕ is calculated, with an example. But we shouldn't rest satisfied until we see, first, that this procedure *must always work*, and second, that the outcome is *uniquely determined by the truth-table definitions.* Now there are infinitely many formulas ϕ to be considered. Hence we have no hope of setting out all the possibilities on a page; we need to invoke some abstract principle to see why the thesis is true.

There is no doubt what principle has to be invoked. It is the principle of *induction on the natural numbers*, otherwise called *mathematical induction.* This principle says the following:

(19) Suppose that the number 0 has a certain property, and suppose also that whenever all numbers from 0 to n inclusive have the property, $n+1$ must also have the property. Then all natural numbers from 0 upwards have the property.

This principle can be put in several forms; the form above is called *course-of-values induction.* (See Appendix B below.) For the moment we shall only be using one or two highly specific instances of it, where the property in question is a mechanically checkable property of arrays of symbols. Several writers have maintained that one knows the truth of any such instance of (19) by a kind of inspection (*Anschauung*). (See for example [Herbrand, 1930, Introduction] and [Hilbert, 1923]. There is a discussion of the point in [Steiner, 1975].)

Essentially what we have to do is to tie a number n to each formula ϕ, calling n the *complexity* of ϕ, so that we can then use induction to prove:

(20) For each number n from 0 upwards, the thesis stated at the beginning of this section is true for all formulas of complexity n.

There are several ways of carrying this through, but they all rest on the same idea, namely this: *all formulas are generated from atomic formulas in a finite number of steps and in a unique way; therefore each formula can be assigned a complexity which is greater than the complexities assigned to any formulas that went into the making of it.* It was Emil Post, one of the founders of formal language theory, who first showed the importance of this idea in his paper on truth-tables:

(21) "It is desirable in what follows to have before us the vision of the totality of these [formulas] streaming out from the unmodified [sentence letters] through forms of ever-growing complexity ..." (Post [1921], p. 266 of van Heijenoort [1967]).

For an exact definition of formulas and their complexities, we need to say precisely what sentence letters we are using. But it would be a pity to lumber ourselves with a set of letters that was inconvenient for some future purposes. So we adopt a compromise. Let X be any set of symbols to be used as sentence letters. Then we shall define the *propositional language of similarity type* X, in symbols $\mathrm{L}(X)$. The set X is not fixed in advance; but as soon as it is fixed, the definition of $\mathrm{L}(X)$ becomes completely precise. This is the usual modern practice.

The notions 'formula of similarity type X' (we say 'formula' for short) and 'complexity of a formula' are defined as follows.

1. Every symbol in X is a formula of complexity 0. $\bot$ is a formula of complexity 1.

2. If ϕ and ψ are formulas of complexities m and n respectively, then $\neg\phi$ is a formula with complexity $m+1$, and $(\phi \wedge \psi)$, $(\phi \vee \psi)$, $(\phi \to \psi)$ and $(\phi \leftrightarrow \psi)$ are formulas of complexity $m+n+1$.

3. Nothing is a formula except as required by (1) and (2).

For definiteness the *language of similarity type* $X, \mathrm{L}(X)$, can be defined as the ordered pair $\langle X, F\rangle$ where F is the set of all formulas of similarity type X. A *propositional language* is a language $\mathrm{L}(X)$ where X is a set of symbols; the *formulas of* $\mathrm{L}(X)$ are the formulas of similarity type X.

Frege would have asked: How do we know there is a unique notion 'formula of similarity type X' with the properties (1)–(3)? A full answer to this question lies in the theory of inductive definitions; cf. Appendix B below. But for the present it will be enough to note that by (1) and (2), every formation tree has a formula as its bottom line, and conversely by (3) every formula is the bottom line of a formation tree. We can prove rigorously by induction that if a formula has complexity n by definition (1)–(3) then it can't also have complexity m where $m \neq n$. This is actually not trivial. It depends on showing that the main connective in a compound formula is uniquely determined, and—ignoring $\neg$ and $\bot$ for simplicity—we can do that by showing that the main connective is the only truth-functor occurrence which has one more '(' than ')' to the left of it. (Cf. [Kleene, 1952, pp. 21ff].) The proof shows at the same time that every formula has a unique formation tree.

The atomic formulas are those which have complexity 0. A formula is called *basic* if it is either atomic or the negation of an atomic formula.

Now that the language has been adequately formulated, we come back to truth-tables. Let L be a propositional language with similarity type X. Then we define an L-*structure* to be a function from the set X to the set $\{T, F\}$ of truth-values. (Set-theoretic notions such as 'function' are defined in Appendix C below, or in any elementary textbook of set theory.) So an L-structure assigns a truth-value to each sentence letter of L. For each sentence letter ϕ we write $I_{\mathfrak{A}}(\phi)$ for the truth-value assigned to ϕ by the L-structure $\mathfrak{A}$. In a truth-table where the sentence letters of L are listed at top left, each row on the left will describe an L-structure, and every L-structure corresponds to just one row of the table.

Now we shall define when a formula ϕ of L is *true in* an L-structure $\mathfrak{A}$, or in symbols

(22) $\mathfrak{A} \models \phi$.

The definition of (22) will be by induction of the complexity of ϕ. This means that when ϕ has low complexity, the truth or falsity of (22) will be determined outright; when ϕ has higher complexity the truth of (22) depends in an unambiguous way on the truth of statements '$\mathfrak{A} \models \psi$' for formulas ψ of lower complexity than ϕ. (Cf. Appendix B.) We can prove by induction on the natural numbers that this definition determines exactly when (22) is true, and in fact that the truth or otherwise of (22) can be calculated mechanically once we know what $\mathfrak{A}$ and ϕ are. The definition is as follows:

(23) For each sentence letter $\phi, \mathfrak{A} \models \phi$ iff $I_{\mathfrak{A}}(\phi) = T$.
It is false that $\mathfrak{A} \models \bot$.
For all formulas ϕ, ψ of L,
$\mathfrak{A} \models \neg\phi$ if it is not true that $\mathfrak{A} \models \phi$;
$\mathfrak{A} \models (\phi \wedge \psi)$ iff $\mathfrak{A} \models \varphi$ and $\mathfrak{A} \models \psi$;
$\mathfrak{A} \models (\varphi \vee \psi)$ iff either $\mathfrak{A} \models \phi$ or $\mathfrak{A} \models \psi$ or both;
$\mathfrak{A} \models (\phi \to \psi)$ iff not: $\mathfrak{A} \models \phi$ but not $\mathfrak{A} \models \psi$.
$\mathfrak{A} \models (\phi \leftrightarrow \psi)$ iff either $\mathfrak{A} \models \phi$ and $\mathfrak{A} \models \psi$, or neither $\mathfrak{A} \models \phi$ nor $\mathfrak{A} \models \psi$.

Definition (23) is known as the *truth definition* for the language L. The statement '$\mathfrak{A} \models \phi$' is sometimes read as: $\mathfrak{A}$ is a *model of* ϕ.

The reader can verify that (23) matches the truth-table definitions of the truth-functors, in the following sense. The left-hand part of any row of a truth-table for ϕ describes an L-structure $\mathfrak{A}$ (for some appropriate language L). The truth-table gives ϕ the value T in this row if and only if $\mathfrak{A} \models \phi$; moreover the steps by which we calculated this value for ϕ in the table exactly match the steps by which the definition (23) above determines whether $\mathfrak{A} \models \phi$. In this sense, and only in this sense, (23) is a correct 'definition of truth for L'. Nobody claims that (23) explains what is meant by the word 'true'.

I should mention a useful piece of notation. We can write $||\phi||_{\mathfrak{A}}$ for the truth-value assigned to the formula ϕ by the structure $\mathfrak{A}$. Then $||\phi||_{\mathfrak{A}}$ can be defined in terms of $\models$ by:

(24) $$||\phi||_{\mathfrak{A}} = \begin{cases} T & \text{if } \mathfrak{A} \models \phi, \\ F & \text{otherwise.} \end{cases}$$

Some writers prefer to define $||\ ||_{\mathfrak{A}}$ directly, and then $\models$ in terms of $||\ ||_{\mathfrak{A}}$. If we write 1 for T and 0 for F, an inductive definition of $||\ ||_{\mathfrak{A}}$ will contain clauses such as

(25) $$||\neg\phi||_{\mathfrak{A}} = 1 - ||\phi||_{\mathfrak{A}}; \quad ||(\phi \vee \psi)||_{\mathfrak{A}} = \max\{||\phi||_{\mathfrak{A}}, ||\psi||_{\mathfrak{A}}\}.$$

4 SOME POINTS OF NOTATION

In Section 3 we put the truth-table method onto a more solid footing. We extended it a little too, because we made no assumption that the language L had just finitely many sentence letters. The original purpose of the exercise was to prove valid argument schemas, and we can now redefine these in sharper terms too.

Let L be a fixed propositional language and $\phi_1, \ldots, \phi_n, \psi$ any formulas of L. Then the statement

(26) $$\phi_1, \ldots, \phi_n \models \psi$$

will mean: for every L-structure $\mathfrak{A}$, if $\mathfrak{A} \models \phi_1$ and $\ldots \mathfrak{A} \models \phi_n$, then $\mathfrak{A} \models \psi$. We allow n to be zero; thus

(27) $$\models \psi$$

means that for every L-structure $\mathfrak{A}, \mathfrak{A} \models \psi$. To say that (26) is false, we write

(28) $$\phi_1, \ldots, \phi_n \not\models \psi.$$

Note that (26)–(28) are statements about formulas of L and not themselves formulas of L.

It is a misfortune that custom requires us to use the same symbol $\models$ both in '$\mathfrak{A} \models \phi$' (cf. (22) above) and in '$\phi_1, \ldots, \phi_n \models \psi$'. It means quite different things in the two cases. But one can always see which is meant, because in the first case a structure $\mathfrak{A}$ is mentioned immediately to the left of $\models$, and in the second usage $\models$ follows either a formula or an empty space. $\models$ can be pronounced 'double turnstile' or 'semantic turnstile', to contrast it with the symbol $\vdash$ ('turnstile' or 'syntactic turnstile') which occurs in the study of formal proof calculi (cf. Section 7 below).

The point of definition (26) should be clear. It says in effect that if we make any consistent replacement of the sentence letters by sentences of English, then in any situation where the sentences resulting from $\phi_1, \ldots, \phi_n$ are true, the sentence got from ψ will be true too. In short (26) says that

(29) $\phi_1, \ldots, \phi_n$. Therefore ψ.

is a valid argument schema. What's more, it says it without mentioning either English sentences or possible situations. Statements of form (26) or (27) are called *sequents* (= 'things that follow' in Latin). When (26) is true, $\phi_1, \ldots, \phi_n$ are said to *logically imply* ψ. When (27) is true, ψ is said to be a *tautology*; for a language with a finite number of sentence letters, this means that the truth-table of ψ has T all the way down its main column. Some elementary texts give long lists of tautologies (e.g. Kalish and Montague [1964, pp. 80– 84]).

While we are introducing notation, let me mention some useful abbreviations. Too many brackets can make a formula hard to read. So we shall agree that when naming formulas we can leave out some of the brackets. First, we can leave off the brackets at the two ends of an occurrence of $(\phi \wedge \psi)$ or $(\phi \vee \psi)$ provided that the only truth-functor which occurs immediately outside them is either $\rightarrow$ or $\leftrightarrow$. For example we can abbreviate

(30) $(p \leftrightarrow (q \wedge r))$ and $((p \wedge q) \rightarrow (r \vee s))$

to

(31) $(p \leftrightarrow q \wedge r)$ and $(p \wedge q \rightarrow r \vee s)$

respectively; but we can *not* abbreviate

(32) $(\neg(p \wedge q) \rightarrow r)$ and $((p \leftrightarrow q) \wedge r)$

to

(33) $(\neg p \wedge q \rightarrow r)$ and $(p \leftrightarrow q \wedge r)$

respectively.

Second, we can leave off brackets at the ends of a formula. So the formulas in (31) can also be written

(34) $p \leftrightarrow q \wedge r$ and $p \wedge q \rightarrow r \vee s$

respectively.

Third, if we have a string of $\wedge$'s with their associated brackets bunched up towards the left end of the formula, as in

(35) $(((q \wedge r) \wedge s) \wedge t)$,

then we can omit all but the outermost brackets:

(36) $(q \wedge r \wedge s \wedge t)$.

Formula (36) is called a *conjunction* whose *conjuncts* are q, r, s, t. Likewise we can abbreviate $(((q \vee r) \vee s) \vee t)$ to the *disjunction* $(q \vee r \vee s \vee t)$ with *disjuncts* q, r, s, t. (But the corresponding move with $\rightarrow$ or $\leftrightarrow$ is not allowed.)

All these conventions can be applied together, as when we write

(37) $p \wedge q \wedge r \rightarrow s$

for

(38) $(((p \wedge q) \wedge r) \rightarrow s)$.

When only these abbreviations are used, it is always possible to work out exactly which brackets have been omitted, so that there is no loss of information.

Jan Łukasiewicz pointed out that if we always write connectives to the left of the formulas they connect, then there is no need for any brackets at all. In this style the second formula of (30) could be written

(39) $\rightarrow \wedge pq \vee rs$, or in Łukasiewicz's notation C$\mathrm{K}pq\mathrm{A}rs$.

Prior [1962] uses Łukasiewicz's notation throughout.

Note that the abbreviations described above only affect the way we talk about formulas of L—the formulas themselves remain untouched. The definition of 'formula of similarity type X' given in Section 3 stands without alteration. Some early writers were rather carefree about this point, making it difficult to follow what language L they really had in mind. If anybody wants to do calculations *in* L but still take advantage of our abbreviations, there is an easy way he can do it. He simply writes down abbreviated *names* of formulas instead of the formulas themselves. In other words, he works always in the metalanguage and never in the object language. This cheap trick will allow him the best of both worlds: a rigorously defined language and a relaxed and generous notation. Practising logicians do it all the time.

5 PROPERTIES OF $\models$

This section gathers up some properties of $\models$ which can be proved directly from the definitions in Sections 3 and 4 above. They are rather a ragbag, but there are some common themes.

THEOREM 1. *If $\mathfrak{A}$ and $\mathfrak{B}$ are structures which assign the same truth-values as each other to each sentence letter occurring in ϕ, then $\mathfrak{A} \models \phi$ iff $\mathfrak{B} \models \phi$.*

This is obvious from (23), but it can also be proved rigorously by induction on the complexity of ϕ. The most important consequence of Theorem 1 is:

THEOREM 2. *The truth of the sequent '$\phi_1, \ldots, \phi_n \models \psi$' doesn't depend on what language* L *the formulas $\phi_1, \ldots, \phi_n$ and ψ come from.*

In other words, although the definition of '$\phi_1, \ldots, \phi_n \models \psi$' was stated in terms of one language L containing $\phi_1, \ldots, \phi_n$ and ψ, any two such languages would give the same outcome. At first sight Theorem 2 seems a reasonable property to expect of any decent notion of entailment. But in other logics, notions of entailment which violate Theorem 2 have sometimes been proposed. (There is an example in Dunn and Belnap [1968], and another in Section 15 below.)

The next result turns all problems about sequents into problems about tautologies.

THEOREM 3 (Deduction Theorem). *$\phi_1, \ldots, \phi_n \models \psi$ if and only if $\phi_1, \ldots, \phi_{n-1} \models \phi_n \to \psi$.*

Theorem 3 moves formulas to the right of $\models$. It has a twin that does the opposite:

THEOREM 4. *$\phi_1, \ldots, \phi_n \models \psi$ iff $\phi_1, \ldots, \phi_n, \neg\psi \models \bot$.*

We say that the formula ϕ is *logically equivalent* to the formula ψ if $\phi \models \psi$ and $\psi \models \phi$. This is equivalent to saying that $\models \phi \leftrightarrow \psi$. Intuitively speaking, logically equivalent formulas are formulas which behave in exactly the same way inside arguments. Theorem 5 makes this more precise:

THEOREM 5. *If $\phi_1, \ldots, \phi_n \models \psi$, and we take an occurrence of a formula χ inside one of $\phi_1, \ldots, \phi_n, \psi$ and replace it by an occurrence of a formula which is logically equivalent to χ, then the resulting sequent holds too.*

For example, $\neg p \vee q$ is logically equivalent to $p \to q$ (as truth-tables will confirm). Also we can easily check that

(40) $r \to (\neg p \vee q), p \models r \to q.$

Then Theorem 5 tells us that the following sequent holds too:

(41) $r \to (p \to q), p \models r \to q.$

An interesting consequence of Theorem 5 is:

THEOREM 6. *Every formula ϕ is logically equivalent to a formula which uses the same sentence letters as ϕ, but no truth-functors except $\bot, \neg$ and $\to$.*

Proof. Truth-tables will quickly show that

(42) $\psi \wedge \chi$ is logically equivalent to $\neg(\psi \to \neg\chi)$,
$\psi \vee \chi$ is logically equivalent to $(\neg\psi \to \chi)$, and
$\psi \leftrightarrow \chi$ is logically equivalent to $\neg((\psi \to \chi) \to \neg(\chi \to \psi))$.

But then by Theorem 5, if we replace a part of ϕ of form $(\psi \wedge \chi)$ by $\neg(\psi \to \neg\chi)$, the resulting formula will be logically equivalent to ϕ. By replacements of this kind we can eliminate in turn all the occurrences of $\wedge, \vee$ and $\leftrightarrow$ in ϕ, and be left with a formula which is logically equivalent to ϕ. This proves Theorem 6. Noting that

(43) $\neg\phi$ is logically equivalent to $\phi \to \bot$,

we can eliminate $\neg$ too, at the cost of introducing some more occurrences of $\bot$. ■

An argument just like the proof of Theorem 6 shows that every formula is logically equivalent to one whose only truth-functors are $\neg$ and $\wedge$, and to one whose only truth-functors are $\neg$ and $\vee$. But there are some limits to this style of reduction: there is no way of eliminating $\neg$ and $\bot$ in favour of $\wedge, \vee, \to$ and $\leftrightarrow$.

The next result is a useful theorem of Post [1921]. In Section 2 we found a truth-table for each formula. Now we go the opposite way and find a formula for each truth-table.

THEOREM 7. *Let P be a truth-table which writes either T or F against each possible assignment of truth-values to the sentence letters $p_1, \ldots, p_n$. Then P is the truth-table of some formula using no sentence letters apart from $p_1, \ldots, p_n$.*

Proof. I sketch the proof. Consider the jth row of the table, and write ϕ_j for the formula $p'_1 \wedge \cdots \wedge p'_n$, where each p'_i is p_i if the jth row makes p_i true, and $\neg p_i$ if the jth row makes p_i false. Then ϕ_j is a formula which is true at just the jth row of the table. Suppose the rows to which the table gives the value T are rows $j_1, \ldots, j_k$. Then take ϕ to be $\phi_{j_1} \vee \cdots \vee \phi_{j_k}$. If the table has F all the way down, take ϕ to be $\bot$. Then P is the truth-table of ϕ. ■

Theorem 7 says in effect that we could never get a more expressive logic by inventing new truth-functors. Anything we could say with the new truth-functors could also be said using the ones we already have.

A formula is said to be in *disjunctive normal form* if it is either $\bot$ or a disjunction of conjunctions of basic formulas (basic = atomic or negated atomic). The proof of Theorem 7 actually shows that P is the truth-table of some formula in disjunctive normal form. Suppose now that we take any formula ψ, work out its truth-table P, and find a formula ϕ in disjunctive normal form with truth-table P. Then ψ and ϕ are logically equivalent, because they have the same truth-table. So we have proved:

THEOREM 8. *Every formula is logically equivalent to a formula in disjunctive normal form.*

One can also show that every formula is logically equivalent to one in *conjunctive normal form*, i.e. either $\neg\bot$ or a conjunction of disjunctions of basic formulas.

LEMMA 9 (Craig's Interpolation Lemma for propositional logic). *If $\psi \vDash \chi$ then there exists a formula ϕ such that $\psi \vDash \phi$ and $\phi \vDash \chi$, and every sentence letter which occurs in ϕ occurs both in ψ and in χ.*

Proof. Let L be the language whose sentence letters are those which occur both in ψ and in χ, and L^+ the language whose sentence letters are those in either ψ or χ. Write out a truth-table for the letters in L, putting T against a row if and only if the assignment of truth-values in that row can be expanded to form a model of ψ. By Theorem 7, this table is the truth-table of some formula ϕ of L. Now we show $\phi \vDash \chi$. Let $\mathfrak{A}$ be any L^+-structure such that $\mathfrak{A} \vDash \phi$. Let $\mathfrak{C}$ be the L-structure which agrees with $\mathfrak{A}$ on all letters in L. Then $\mathfrak{C} \vDash \phi$ by Theorem 1. By the definition of ϕ it follows that some model $\mathfrak{B}$ of ψ agrees with $\mathfrak{C}$ on all letters in L. Now we can put together an L^+-structure $\mathfrak{D}$ which agrees with $\mathfrak{B}$ on all letters occurring in ψ, and with $\mathfrak{A}$ on all letters occurring in χ. (The overlap was L, but $\mathfrak{A}$ and $\mathfrak{B}$ both agree with $\mathfrak{C}$ and hence with each other on all letters in L.) Then $\mathfrak{D} \vDash \psi$ and hence $\mathfrak{D} \vDash \chi$ since $\psi \vDash \chi$; but then $\mathfrak{A} \vDash \chi$ too. The proof that $\psi \vDash \phi$ is easier and I leave it to the reader. ■

Craig's Lemma is the most recent fundamental discovery in propositional logic. It is easy to state and to prove, but it was first published over a hundred years after propositional logic was invented [Craig, 1957a]. The corresponding lemma holds for full first-order logic too; this is much harder to prove. (See Lemma 32 below.)

Most of the topics in this section are taken further in Hilbert and Bernays [1934], Kleene [1952], Rasiowa and Sikorski [1963] and Bell and Machover [1977].

6 DECISION METHODS

Propositional logic owes much of its flavour to the fact that all interesting problems within it can be solved by scribbled calculations on the back of an envelope. If somebody presents you with formulas $\phi_1, \ldots, \phi_n$ and ψ, and asks you whether $\phi_1, \ldots, \phi_n$ logically imply ψ, then you can calculate the answer as follows. First choose a language L whose sentence letters are just those which occur in the formulas $\phi_1, \ldots, \phi_n, \psi$. If L has k sentence letters then there are just 2^k different L-structures. For each such structure you can check in a finite time whether it is a model of ϕ_1 and ... and ϕ_n but not of ψ. If you find an L-structure with these properties, then $\phi_1, \ldots, \phi_n$ don't logically imply ψ; if you don't, they do. This is the truth-table *decision method for logical implication.*

The question I want to consider next is whether this decision method can be improved. This is not a precise question. Some alternatives to truth-tables are very fast for very short sequents but get quite unwieldy for long ones. Other alternatives grow steadily more efficient as we progress to longer sequents. Some methods are easy to run on a computer but messy on paper; some are as handy one way as another.

Let me sketch one alternative to truth-tables. An example will show the gist. We want to determine whether the following sequent holds.

(44) $p \wedge q, \neg(p \wedge r) \models \neg r.$

By Theorem 4, (44) holds if and only if

(45) $p \wedge q, \neg(p \wedge r), \neg\neg r \models \bot.$

Now (45) says that any structure in which all the formulas on the left of $\models$ are true is a model of $\bot$. But $\bot$ has no models; so (45) says *there is no model of* $p \wedge q, \neg(p \wedge r)$ *and* $\neg\neg r$ *simultaneously.* We try to refute this by constructing such a model. At each stage we ask: what must be true in the model for it to be a model of these sentences? For example, $\neg\neg r$ is true in a structure $\mathfrak{A}$ if and only if r is true in $\mathfrak{A}$; so we can simplify (45) by replacing $\neg\neg r$ by r:

(46)
$$\begin{array}{l} p \wedge q, \quad \neg(p \wedge r), \quad \neg\neg r \models \bot \\ \qquad\qquad\quad | \\ p \wedge q, \quad \neg(p \wedge r), \quad r \models \bot. \end{array}$$

Likewise a structure is a model of $p \wedge q$ if and only if it is both a model of p and a model of q; so we can replace $p \wedge q$ by the two formulas p and q:

(47)
$$\begin{array}{c} p \wedge q, \neg(p \wedge r), \neg\neg r \models \bot \\ p \wedge q, \neg(p \wedge r), r \models \bot \\ | \\ p, q, \neg(p \wedge r), r \models \bot \end{array}$$

Now there are just two ways of making $\neg(p \wedge r)$ true, namely to make $\neg p$ true and to make $\neg r$ true. (Of course these ways are not mutually exclusive.) So in our attempt to refute (45) we have two possible options to try, and the diagram accordingly branches in two directions:

(48)

$$p \wedge q, \neg(p \wedge r), \neg\neg r \models \bot$$
$$p \wedge q, \neg(p \wedge r), r \models \bot$$
$$p, q, \neg(p \wedge r), r \models \bot$$
$$p, q, \neg p, r \models \bot \qquad\qquad p, q, \neg r, r \models \bot$$

But there is no chance of having both p and $\neg p$ true in the same structure. So the left-hand fork is a non-starter, and we block it off with a line. Likewise the right-hand fork expects a structure in which $\neg r$ and r are both true, so it must be blocked off:

(49)

$$p \wedge q, \neg(p \wedge r), \neg\neg r \models \bot$$
$$p \wedge q, \neg(p \wedge r), r \models \bot$$
$$p, q, \neg(p \wedge r), r \models \bot$$
$$\underline{p, q, \neg p, r \models \bot} \qquad\qquad \underline{p, q, \neg r, r \models \bot}$$

Since every possibility has been explored and closed off, we conclude that there is no possible way of refuting (45), and so (45) is correct.

What happens if we apply the same technique to an incorrect sequent? Here is an example:

(50) $p \vee \neg(q \to r), q \to r \models q.$

I leave it to the reader to check the reasons for the steps below—he should note that $q \to r$ is true if and only if either $\neg q$ is true or r is true:

(51)

$$p \vee \neg(q \to r), q \to r, \neg q \models \bot$$
$$p, q \to r, \neg q \models \bot \qquad\qquad \underline{\neg(q \to r), q \to r, \neg q \models \bot}$$
$$p, \neg q, \neg q \models \bot \qquad\qquad p, r, \neg q \models \bot$$

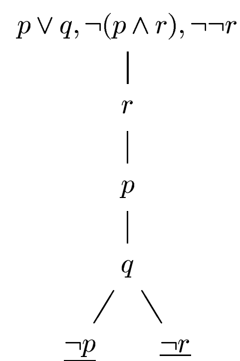
Here two branches remain open, and since all the formulas in them have been decomposed into atomic formulas or negations of atomic formulas, there is nothing more we can do with them. In every such case it turns out that each open branch describes a structure which refutes the original sequent. For example, take the leftmost branch in (51). The formulas on the left side of the bottom sequent describe a structure $\mathfrak{A}$ in which p is true and q is false. The sequent says nothing about r, so we can make an arbitrary choice: let r be false in $\mathfrak{A}$. Then $\mathfrak{A}$ is a structure in which the two formulas on the left in (50) are true but that on the right is false.

This method always leads in a finite time *either* to a tree diagram with all branches closed off, in which case the beginning sequent was correct; *or* to a diagram in which at least one branch remains resolutely open, in which case this branch describes a structure which shows that the sequent was incorrect.

Diagrams constructed along the lines of (49) or (51) above are known as *semantic tableaux*. They were first invented, upside-down and with a different explanation, by Gentzen [1934]. The explanation given above is from Beth [1955] and Hintikka [1955].

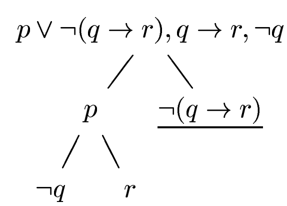
We can cut out a lot of unnecessary writing by omitting the '$\models \bot$' at the end of each sequent. Also in all sequents below the top one, we need only write the new formulas. In this abbreviated style the diagrams are called *truth-trees*. Written as truth-trees, (49) looks like this:

(52)



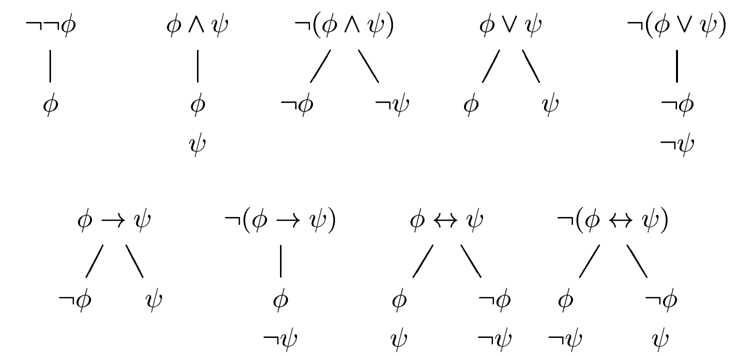
and (51) becomes

(53)



The rules for breaking down formulas in truth-trees can be worked out straight from the truth-table definitions of the truth-functors, but for the reader's convenience I list them:

(54)



One is allowed to close off a branch as soon as either $\bot$ or any outright contradiction $\phi, \neg\phi$ appears among the formulas in a branch. (Truth-trees are used in Jeffrey [1967]; see [Smullyan, 1968; Bell and Machover, 1977] for mathematical analyses.) Truth-trees are one dialect of semantic tableaux. Here is another. We shall understand the generalised sequent

(55) $\phi_1, \ldots, \phi_n \vDash \psi_1, \ldots, \psi_m$

to mean that there is no structure which makes $\phi_1, \ldots, \phi_n$ all true and $\psi_1, \ldots, \psi_m$ all false. A structure in which $\phi_1, \ldots, \phi_n$ are true and $\psi_1, \ldots, \psi_m$ are false is called a *counterexample* to (55). When there is only one formula to the right of $\vDash$, (55) means just the same as our previous sequents (26).

Generalised sequents have the following two symmetrical properties:

(56) $\phi_1, \ldots, \phi_n, \neg\chi \vDash \psi_1, \ldots, \psi_m$ iff $\phi_1, \ldots, \phi_n \vDash \psi_1, \ldots, \psi_m, \chi$.

(57) $\phi_1, \ldots, \phi_n, \vDash \psi_1, \ldots, \psi_m, \neg\chi$ iff $\phi_1, \ldots, \phi_n, \chi \vDash \psi_1, \ldots, \psi_m$.

Suppose now that we construct semantic tableaux as first described above, but using *generalised* sequents instead of sequents. The effect of (56) and (57) is that we handle $\neg$ *by itself*; as (54) shows, our previous tableaux could only tackle $\neg$ two at a time or in combination with another truth-functor.

Using generalised sequents, a proof of (44) goes as follows:

(58)
$$p \wedge q, \neg(p \wedge r) \models \neg r$$
(i)
$$p \wedge q, \neg(p \wedge r), r \models$$
(ii)
$$p \wedge q, r \models p \wedge r$$
(iii)
$$p, q, r \models p \wedge r$$
(iv)
$$\underline{p, q, r \models p} \qquad \underline{p, q, r \models r}$$

Steps (i) and (ii) are by (57) and (56) respectively. Step (iv) is justified as follows. We are trying to build a structure in which p, q and r are true but $p \wedge r$ is false, as a counterexample to the sequent '$p, q, r \vDash p \wedge r$'. By the truth-table for $\wedge$, it is necessary and sufficient to build *either* a structure in which p, q, r are true and p is false, *or* a structure in which p, q, r are true and r is false. We can close off under the bottom left sequent '$p, q, r \vDash p$' because a formula p occurs both on the right and on the left of $\vDash$, so that in a counterexample it would have to be both false and true, which is impossible. Likewise at bottom right.

Proofs with generalised sequents are virtually identical with the *cut-free sequent proofs* of [Gentzen, 1934], except that he wrote them upside down. Beth [1955; 1962] used them as a method for testing sequents. He wrote them in a form where, after the first sequent, one only needs to mention the new formulas.

Quine [1950] presents another quite fast decision method which he calls *fell swoop* (to be contrasted with the 'full sweep' of truth-tables).

I turn to the question how fast a decision method of testing sequents can be in the long run, i.e. as the number and lengths of the formulas increase. At the time of writing, this is one of the major unsolved problems of computation theory. A function $p(n)$ of the number n is said to be *polynomial* if it is calculated from n and some other fixed numbers by adding, subtracting and multiplying. (So for example $n^2 + 3$ and $2n^3 - n$ are polynomial functions of n but $3^n, n!$ and $1/(n^2 + 1)$ are not.) It is not known whether there exist a decision method M for sequents of propositional logic, and a polynomial function $p(n)$, such that for every sequent S, if n is the number of symbols in S then M can determine in less than $p(n)$ steps whether or not S is correct. If the answer is Yes there are such M and $p(n)$, then we say that the decision problem for propositional logic is *solvable in polynomial time.* Cook [1971] showed that a large number of other interesting computational problems will be solvable in polynomial time if this one is. (See [Garey and Johnson, 1979].) I have the impression that everybody working in the field expects the answer to be No. This would mean in effect that for longer sequents the problem is too hard to be solved efficiently by a deterministic computer.

7 FORMAL PROOF CALCULI

During the first third of this century, a good deal of effort was put into constructing various formal proof calculi for logic. The purpose of this work was to reduce reasoning—or at least a sizeable part of mathematical reasoning—to precise mechanical rules. I should explain at once what a *formal proof calculus* (or *formal system*) is.

A formal proof calculus, call it Σ, is a device for proving sequents in a language L. First, Σ gives us a set of rules for writing down arrays of symbols on a page. An array which is written down according to the rules is called a *formal proof* in Σ. The rules must be such that one can check by inspection and calculation whether or not an array is a formal proof. Second, the calculus contains a rule to tell us how we can mechanically work out what are the *premises* and the *conclusion* of each formal proof.

We write

(59) $\phi_1, \ldots, \phi_n \vdash_\Sigma \psi$ or more briefly $\phi_1, \ldots, \phi_n \vdash \psi$

to mean that there is a formal proof in the calculus Σ whose premises all appear in the list $\phi_1, \ldots, \phi_n$, and whose conclusion is ψ. Some other ways of expressing (59) are:

> '$\phi_1, \ldots, \phi_n \vdash \psi$' is a *derivable sequent* of Σ;
> ψ is *deducible from* $\phi_1, \ldots, \phi_n$ in Σ;
> $\phi_1, \ldots, \phi_n$ *yield* ψ in Σ.

We call ψ a *derivable formula* of Σ if there is a formal proof in Σ with conclusion ψ and no premises. The symbol $\vdash$ is called *turnstile* or *syntactic turnstile*.

We say that the calculus Σ is:

> *sound* if $\phi_1, \ldots, \phi_n \vdash_\Sigma$ implies $\phi_1, \ldots, \phi_n \models \psi$
> *strongly complete* if $\phi_1, \ldots, \phi_n \models \psi$ implies $\phi_1, \ldots, \phi_n \vdash_\Sigma \psi$,
> *weakly complete* if $\models \psi$ implies $\vdash_\Sigma \psi$,

where $\phi_1, \ldots, \phi_n, \psi$ range over the formulas of L. These definitions also make sense when $\models$ is defined in terms of other logics, not necessarily first-order. In this chapter 'complete' will always mean 'strongly complete'.

The formal proofs in a calculus Σ are in general meaningless arrays of symbols. They need not be genuine proofs, that is, demonstrations that something is the case. But if we know that Σ is sound, then the fact that a certain sequent is derivable in Σ will prove that the corresponding sequent with $\models$ is correct. In some proof calculi the formal proofs are made to look as much as possible like intuitively correct reasoning, so that soundness can be checked easily.

We already have the makings of one formal proof calculus in Section 6 above: the *cut-free sequent proofs* using generalised sequents. As proofs, these are usually written the other way up, with $\vdash$ in place of $\models$, and with horizontal lines separating the sequents. Also there is no need to put in the lines which mark the branches that are closed, because every branch is closed.

For example, here is a cut-free sequent proof of the sequent '$p \wedge q, \neg(p \wedge r) \vdash \neg r$'; compare it with (58):

$$\text{(60)} \qquad \frac{\dfrac{\dfrac{\dfrac{\dfrac{p \vdash p}{p,q,r \vdash p} \qquad \dfrac{r \vdash r}{p,q,r \vdash r}}{p,q,r \vdash p \wedge r}}{p \wedge q, r \vdash p \wedge r}}{p \wedge q, \neg(p \wedge r), r \vdash}}{p \wedge q, \neg(p \wedge r) \vdash \neg r}$$

To justify this proof we would show, working upwards from the bottom, that if there is a counterexample to the bottom sequent then at least one of the top sequents has a counterexample, which is impossible. Or equivalently, we could start by noting that the top sequents are correct, and then work *down* the tree, showing that each of the sequents must also be correct. By this kind of argument we can show that the cut-free sequent calculus is sound.

To prove that the calculus is complete, we borrow another argument from Section 6 above. Assuming that a sequent S is not derivable, we have to prove that it is not correct. To do this, we try to construct a cut-free sequent proof, working upwards from S. After a finite number of steps we shall have broken down the formulas as much as possible, but the resulting diagram can't be a proof of S because we assumed there isn't one. So at least one branch must still be 'open' in the sense that it hasn't revealed any immediate contradiction. Let B be such a branch. Let B_L be the set of all formulas which occur to the left of $\vdash$ in some generalised sequent in B, and let B_R be the same with 'right' for 'left'. We can define a structure $\mathfrak{A}$ by

$$\text{(61)} \quad I_{\mathfrak{A}}(\phi) = \begin{cases} T & \text{if } \phi \text{ is a sentence letter which is in } B_L, \\ F & \text{if } \phi \text{ is a sentence letter not in } B_L. \end{cases}$$

Then we can prove, by induction on the complexity of the formula ψ, that if ψ is any formula in B_L then $\mathfrak{A} \models \psi$, and if ψ is any formula in B_R then $\mathfrak{A} \models \neg\psi$. It follows that $\mathfrak{A}$ is a counterexample to the bottom sequent S, so that S is not correct.

The cut-free sequent calculus itself consists of a set of mechanical rules for constructing proofs, and it could be operated by somebody who had not the least idea what $\vdash$ or any of the other symbols mean. These rules are listed in Sundholm (in Volume 2 of this *Handbook.*).

Gentzen [1934] had another formal proof calculus, known simply as the *sequent calculus*. This was the same as the cut-free sequent calculus, except that it allowed a further rule called the *cut rule* (because it cuts out a formula):

$$(62)\quad \frac{\ldots \vdash ***, \chi \qquad \ldots, \chi \vdash ***}{\ldots \vdash ***}$$

This rule often permits much shorter proofs. Gentzen justified it by showing that any proof which uses the cut rule can be converted into a cut-free proof of the same sequent. This *cut elimination theorem* is easily the best mathematical theorem about proofs. Gentzen himself adapted it to give a proof of the consistency of first-order Peano arithmetic. By analysing Gentzen's argument we can get sharp information about the degree to which different parts of mathematics rely on infinite sets. (Cf. [Schütte, 1977]. Gentzen's results on cut-elimination were closely related to deep but undigested work on quantifier logic which Jacques Herbrand had done before his death in a mountaineering accident at the age of 23; see [Herbrand, 1930] and the Introduction to [Herbrand, 1971].) Further details of Gentzen's sequent calculi, including the intuitionistic versions, are given in [Kleene, 1952, Ch XV] and Sundholm (in Volume 2 of this *Handbook*).

In the same paper, Gentzen [1934] described yet a third formal proof calculus. This is known as the *natural deduction calculus* because proofs in this calculus start with their premises and finish at their conclusions (unlike sequent calculi and semantic tableaux), and all the steps between are intuitively natural (unlike the Hilbert-style calculi to be described below).

A proof in the natural deduction calculus is a tree of formulas, with a single formula at the bottom. The formulas at the tops of the branches are called the *assumptions* of the proof. Some of the assumptions may be *discharged* or *cancelled* by having square brackets [ ] written around them. The *premises* of the proof are its uncancelled assumptions, and the *conclusion* of the proof is the formula at the bottom.

Sundholm (in his chapter in Volume D2 of this *Handbook*) gives the full rules of the natural deduction calculus. Here are a few illustrations. Leaving aside $\neg$ and $\bot$ for the moment, there are two rules for each truth-functor, namely an *introduction* rule and an *elimination* rule. The introduction rule for $\wedge$ is:

$$(63)\quad \frac{\phi \qquad \psi}{\phi \wedge \psi}$$

i.e. from ϕ and ψ deduce $\phi \wedge \psi$. The elimination rule for $\wedge$ comes in a left-hand version and a right-hand version:

(64) $$\frac{\phi \wedge \psi}{\phi} \qquad\qquad \frac{\phi \wedge \psi}{\psi}.$$

The introduction rule for $\to$ says that if we have a proof of ψ from certain assumptions, then we can deduce $\phi \to \psi$ from those assumptions less ϕ:

(65) $$\frac{\begin{array}{c}[\phi]\\ \vdots \\ \psi\end{array}}{\phi \to \psi}$$

The elimination rule for $\to$ is the *modus ponens* of the medievals:

(66) $$\frac{\phi \qquad \phi \to \psi}{\psi}.$$

For example, to prove

(67) $q, p \wedge q \to r \vDash p \to r$

in the natural deduction calculus we write:

(68) $$\cfrac{\cfrac{\cfrac{[p] \quad q}{p \wedge q} \qquad p \wedge q \to r}{r}}{p \to r}$$

Note that the assumption p is discharged at the last step when $p \to r$ is introduced.

The calculus reads $\neg\phi$ as a shorthand for $\phi \to \bot$. So for example, from ϕ and $\neg\phi$ we deduce $\bot$ by (66). There is an elimination rule for $\bot$. It says: given a proof of $\bot$ from certain assumptions, derive ϕ from the same assumptions less $\phi \to \bot$:

(69) $$\frac{\begin{array}{c}[\phi \to \bot]\\ \vdots \\ \bot\end{array}}{\phi}$$

This is a form of *reductio ad absurdum.*

The rule about cancelling assumptions in (65) should be understood as follows. When we make the deduction, we are *allowed* to cancel ϕ wherever it occurs as an assumption. But we are not obliged to; we can cancel some occurrences of ϕ and not others, or we can leave it completely uncancelled. The formula ϕ may not occur as an assumption anyway, in which case we can forget about cancelling it. The same applies to $\phi \to \bot$ in (69). So (69) implies the following weaker rule in which we make no cancellations:

$$(70)\quad \frac{\bot}{\phi}$$

('Anything follows from a contradiction'.) Intuitionist logic accepts (70) but rejects the stronger rule (69) (cf. van Dalen (Volume 7).

Belnap [1962] and Prawitz [1965] have explained the idea behind the natural deduction calculus in an interesting way. For each truth-functor the rules are of two sorts, the introduction rules and the elimination rules. In every case the elimination rules *only allow us to infer from a formula what we had to know in order to introduce the formula.* For example we can remove $\phi \to \psi$ only by rule (66), i.e. by using it to deduce ψ from ϕ; but $\phi \to \psi$ can only be introduced either as an explicit assumption or (by (65)) when we already know that ψ can be deduced from ϕ. (Rule (69) is in a special category. It expresses (1) that everything is deducible from $\bot$, and (2) that for each formula ϕ, at least one of ϕ and $\phi \to \bot$ is true.)

Popper [1946/47, particularly p. 284] rashly claimed that he could define truth-functors just by writing down natural deduction rules for them. Prior [1960] gave a neat example to show that this led to absurdities. He invented the new truth-functor *tonk*, which is defined by the rules

$$(71)\quad \frac{\phi}{\phi\ \mathit{tonk}\ \psi} \qquad \frac{\phi\ \mathit{tonk}\ \psi}{\psi}$$

and then proceeded to infer everything from anything. Belnap [1962] points out that Prior's example works because its introduction and elimination rules fail to match up in the way described above. Popper should at least have imposed a requirement that the rules must match up. (Cf. [Prawitz, 1979], [Tennant, 1978, p. 74ff], and Sundholm (Volume 2).)

Natural deduction calculi, all of them variants of Gentzen's, are given by Anderson and Johnstone [1962], Fitch [1952], Kalish and Montague [1964], Lemmon [1965], Prawitz [1965], Quine [1950], Suppes [1957], Tennant [1978], Thomason [1970] and van Dalen [1980]. Fitch (followed e.g. by Thomason) makes the trees branch to the right. Some versions (e.g. Quine's) disguise the pattern by writing the formulas in a vertical column. So they have to supply some other way of marking which formulas depend on which assumptions; different versions do this in different ways.

Just as a semantic tableau with its branches closed is at heart the same thing as a cut-free sequent proof written upside down, Prawitz [1965] has shown that after removing redundant steps, a natural deduction proof is really the same thing as a cut-free sequent proof written sideways. (See also Zucker [1974].) The relationship becomes clearer if we adapt the natural deduction calculus so as to allow a proof to have several alternative conclusions, just as it has several premises. Details of such calculi have been worked out by Kneale [1956] and more fully by Shoesmith and Smiley [1978].

A proof of $p \vee \neg p$ in Gentzen's natural deduction calculus takes several lines. This is a pity, because formulas of the form $\phi \vee \neg\phi$ are useful halfway steps in proofs of other formulas. So some versions of natural deduction allow us to quote a few tautologies such as $\phi \vee \neg\phi$ whenever we need them in a proof. These tautologies are then called *axioms*. Technically they are formulas deduced from no assumptions, so we draw a line across the top of them, as at top right in (72) below.

If we wanted to undermine the whole idea of natural deduction proofs, we could introduce axioms which replace all the natural deduction rules except modus ponens. For example we can put (63) out of a job by using the axiom $\phi \to (\psi \to \phi \wedge \psi)$. Whenever Gentzen used (63) in a proof, we can replace it by

$$\text{(72)} \qquad \dfrac{\psi \qquad \dfrac{\phi \qquad \overline{\phi \to (\psi \to \phi \wedge \psi)}}{\psi \to \phi \wedge \psi}}{\phi \wedge \psi}$$

using (66) twice. Likewise (64) become redundant if we use the axioms $\phi \wedge \psi \to \phi$ and $\phi \wedge \psi \to \psi$. Rule (65) is a little harder to dislodge, but it can be done, using the axioms $\phi \to (\psi \to \phi)$ and $(\phi \to \psi) \to ((\phi \to (\psi \to \chi)) \to (\phi \to \chi))$.

At the end of these manipulations we have what is called a *Hilbert-style* proof calculus. A Hilbert-style calculus consists of a set of formulas called *axioms*, together with one or two *derivation rules* for getting new formulas out of given ones. To prove $\phi_1, \ldots, \phi_n \vDash \psi$ in such a calculus, we apply the derivation rules as many times as we like to $\phi_1, \ldots, \phi_n$ and the axioms, until they give us ψ.

One Hilbert-style system is described in Appendix A below. Mates [1965] works out another such system in detail. Hilbert-style calculi for propositional logic were given by Frege [1879; 1893], Peirce [1885], Hilbert [1923] and Łukasiewicz (see [Łukasiewicz and Tarski, 1930]). (Cf. Sundholm (Volume 2 of this *Handbook*).)

The typical Hilbert-style calculus is inefficient and barbarously unintuitive. But they do have two merits. The first is that their mechanics are usually very simple to describe—many Hilbert-style calculi for propositional logic have only one derivation rule, namely modus ponens. This makes them suitable for encoding into arithmetic (Section 24 below). The second merit is that we can strengthen or weaken them quite straightforwardly by tampering with the axioms, and this commends them to researchers in non-classical logics.

Soundness for these calculi is usually easy to prove: one shows (a) that the axioms are true in every structure and (b) that the derivation rules never lead from truth to falsehood. One way of proving completeness is to show that every natural deduction proof can be converted into a Hilbert-style proof of the same sequent, as hinted above. (Kleene [1952] Section 77 shows how to convert sequent proofs into Hilbert-style proofs and *vice versa*; see Sundholm (Volume 2 of this *Handbook*).)

Alternatively we can prove their completeness directly, using maximal consistent sets. Since this is a very un-proof-theoretic approach, and this section is already too long, let me promise to come back to the matter at the end of Section 16 below. (Kalmár [1934/5] and Kleene independently found a neat proof of the weak completeness of Hilbert-style calculi, by converting a truth-table into a formal proof; cf. Kleene [1952, p. 132ff] or Mendelson [1987, p. 34].)

II: Predicate Logic

8 BETWEEN PROPOSITIONAL LOGIC AND PREDICATE LOGIC

If we asked a Proof Theorist to explain what it means to say

(73) $\phi_1, \ldots, \phi_n$ logically imply ψ,

where $\phi_1, \ldots, \phi_n$ and ψ are formulas from propositional logic, he would explain that it means this: there is a proof of ψ from $\phi_1, \ldots, \phi_n$ in one of the standard proof calculi. A Model Theorist would prefer to use the definition we gave in Section 4 above, and say that (73) means: whenever $\phi_1, \ldots, \phi_n$ are true in a structure, then ψ is true in that structure too. The Traditional Logician for his part would explain it thus: every argument of the form '$\phi_1, \ldots, \phi_n$. *Therefore* ψ' is valid. There need be no fight between these three honest scholars, because it is elementary to show that (73) is true under any one of these definitions if and only if it is true under any other.

In the next few sections we shall turn from propositional logic to predicate logic, and the correct interpretation of (73) will become more contentious.

When $\phi_1, \ldots, \phi_n$ and ψ are sentences from predicate logic, the Proof Theorist has a definition of (73) which is a straightforward extension of his definition for propositional logic, so he at any rate is happy.

But the Traditional Logician will be in difficulties, because the quantifier expressions of predicate logic have a quite different grammar from all locutions of normal English; so he is hard put to say what would count as an argument of the form '$\phi_1, \ldots, \phi_n$. *Therefore* ψ'. He will be tempted to say that really we should look at sentences whose deep structures (which he may call logical forms) are like the formulas $\phi_1, \ldots, \phi_n, \psi$. This may satisfy him, but it will hardly impress people who know that in the present state of the linguistic art one can find experts to mount convincing arguments for any one of seventeen deep structures for a single sentence. A more objective but admittedly vague option would be for him to say that (73) means that any argument which can be *paraphrased* into this form, using the apparatus of first-order logic, is valid.

But the man in the worst trouble is the Model Theorist. On the surface all is well—he has a good notion of 'structure', which he took over from the algebraists, and he can say just what it means for a formula of predicate logic to be 'true in' a structure. So he can say, just as he did for propositional logic, that (73) means that whenever $\phi_1, \ldots, \phi_n$ are true in a structure, then ψ is true in that structure too. His problems start as soon as he asks himself what a structure really is, and how he knows that they exist.

Structures, as they are presented in any textbook of model theory, are abstract set-theoretic objects. There are uncountably many of them and most of them are infinite. They can't be inspected on a page (like proofs in a formal calculus) or heard at Hyde Park Corner (like valid arguments). True, several writers have claimed that the only structures which exist are those which somebody constructs. (E.g. Putnam [1980, p. 482]: 'Models are ... constructions within our theory itself, and they have names from birth.') Unfortunately this claim is in flat contradiction to about half the major theorems of model theory (such as the Upward Löwenheim–Skolem Theorem, Theorem 14 in Section 17 below).

Anybody who wants to share in present-day model theory has to accept that structures are as disparate and intangible as sets are. One must handle them by set-theoretic principles and not by explicit calculation. Many model theorists have wider horizons even than this. They regard the whole universe V of sets as a structure, and they claim that first-order formulas in the language of set theory are true or false in this structure by just the same criteria as in smaller structures. The axioms of Zermelo–Fraenkel set theory, they claim, are simply true in V.

It is actually a theorem of set theory that a notion of truth adequate to cope with the whole universe of sets *cannot be formalised within set theory.* (We prove this in Section 24 below.) So a model theorist with this wider horizon is strictly not entitled to use formal set-theoretic principles either, and he is forced back onto his intuitive understanding of words like 'true', 'and', 'there is' and so forth. In mathematical practice this causes no problems whatever. The problems arise when one tries to justify what the mathematicians are doing.

In any event it is a major exercise to show that these three interpretations of (73) in predicate logic—or four if we allow the Model Theorist his wider and narrower options—agree with each other. But logicians pride themselves that it can be done. Section 17 will show how.

9 QUANTIFIERS

First-order predicate logic comes from propositional logic by adding the words 'every' and 'some'.

Let me open with some remarks about the meaning of the word 'every'. There is no space here to rebut rival views (Cf. Leblanc (see Volume 2 of this *Handbook*); on substitutional quantification see [Dunn and Belnap, 1968; Kripke, 1976; Stevenson, 1973].) But anybody who puts a significantly different interpretation on 'every' from the one presented below will have to see first-order logic in a different light too.

A person who understands the words 'every', 'Pole', the sentence

(74) Richard is a Catholic.

and the principles of English sentence construction must also understand the sentence

(75) Every Pole is a Catholic.

How?

First, (74) is true if and only if Richard satisfies a certain condition, namely that

(76) <u>He</u> is a Catholic.

I underline the pronoun that stands for whatever does or does not satisfy the condition. Note that the condition expressed by (76) is one which people either satisfy or do not satisfy, regardless of how or whether we can identify them. Understanding the condition is a necessary part of understanding (74). In Michael Dummett's words [1973, p. 517]:

> ... given that we understand a sentence from which a predicate has been formed by omission of certain occurrences of a name, we are capable of recognising what concept that predicate stands for in the sense of knowing what it is for it to be true of or false of any arbitrary object, whether or not the language contains a name for that object.

Second, the truth or otherwise of (75) in a situation depends on what class of Poles is on the agenda. Maybe only Poles at this end of town are under discussion, maybe Poles anywhere in the world; maybe only Poles alive now, maybe Poles for the last hundred years or so. Possibly the speaker was a little vague about which Poles he meant to include. I count the specification of the relevant class of Poles as part of the situation in which (75) has a truth-value. This class of Poles is called the *domain of quantification* for the phrase 'every Pole' in (75). The word 'Pole' is called the *restriction term*, because it restricts us to Poles; any further restrictions on the domain of quantification are called *contextual restrictions*.

So when (75) is used in a context, the word 'Pole' contributes a domain of quantification and the words 'is a Catholic' contribute a condition. The contribution of the word 'Every' is as follows: *In any situation*, (75) *is true iff every individual in the domain of quantification satisfies the condition*.

This analysis applies equally well to other simple sentences containing 'Every', such as:

(77) She ate every flower in the garden.

For (77), the situation must determine what the garden is, and hence what is the class of flowers that were in the garden. This class is the domain of quantification; 'flower in the garden' is the restriction term. The sentence

(78) She ate <u>it</u>.

expresses a condition which things do or do not satisfy, once the situation has determined who 'she' refers to. So in this example the condition varies with the situation. The passage from condition and domain of quantification to truth-value is exactly as before.

The analysis of

(79) Some Pole is a Catholic

(80) She ate some flower (that was) in the garden,

is the same as that of (75), (77) respectively, except at the last step. For (79) or (80) to be true we require that *at least one individual in the domain of quantification satisfies the condition*.

In the light of these analyses we can introduce some notation from first-order logic. In place of the underlined pronoun in (76) and (78) we shall use an *individual variable*, i.e. (usually) a lower-case letter from near the end of the alphabet, possibly with a subscript. Thus:

(81) x is a Catholic.

Generalising (81), we use the phrase *1-place predicate* to mean a string consisting of words and one individual variable (which may be repeated), such that if the variable is understood as a pronoun referring to a certain person or object, then the string becomes a sentence which expresses that the person or object referred to satisfies a certain condition. The condition may depend on the situation into which the sentence is put.

For an example in which a variable occurs twice,

(82) x handed the melon to Schmidt, who gave it back to x.

is a 1-place predicate. It expresses the condition which Braun satisfies if and only Braun handed the melon to Schmidt and Schmidt gave it back to Braun.

To return to (75), 'Every Pole is a Catholic': we have now analysed this sentence into (a) a *quantifier word* 'Every', (b) the restriction term 'Pole', and (c) the predicate 'x is a Catholic'.

The separating out of the predicate (by [Frege, 1879], see also [Mitchell, 1883] and [Peirce, 1883]) was vital for the development of modern logic. Predicates have the grammatical form of sentences, so that they can be combined by truth-functors. For example

(83) (x is a Catholic $\wedge$ x is a philatelist)

is a predicate which is got by conjoining two other predicates with $\wedge$. It expresses the condition which a person satisfies if he is both a Catholic and a philatelist. Incidentally I have seen it suggested that the symbol $\wedge$ must have a different meaning in (83) from its meaning in propositional logic, because in (83) it stands between predicates which do not have truth-values. The answer is that predicates *do* gain truth-values when their variables are either replaced by or interpreted as names. The truth-value gained in this way by the compound predicate (83) is related to the truth-values gained by its two conjuncts in exactly the way the truth-table for $\wedge$ describes.

(A historical aside: Peirce [1885] points out that by separating off the predicate we can combine quantifiers with propositional logic; he says that all attempts to do this were 'more or less complete failures until Mr Mitchell showed how it was to be effected'. Mitchell published in a volume of essays by students of Peirce at Johns Hopkins [Members of the Johns Hopkins University, Boston, 1883]. Christine Ladd's paper in the same volume mentions both Frege's *Begriffschrift* [1879] and Schröder's review of it. It is abundantly clear that nobody in Peirce's group had read either. The same happens today.)

The account of quantifiers given above agrees with what Frege said in his *Funktion und Begriff* [1891] and *Grundgesetze* [1893], except in one point. Frege required that all conditions on possible values of the variable should be stated in the predicate. In other words, he allowed only one domain of quantification, namely absolutely everything. For example, if someone were to say, à propos of Poles in New York, 'Every Pole is a Catholic', Frege would take this to mean that absolutely everything satisfies the condition

(84) If x is a Pole in New York City then x is a Catholic.

If a person were to say

(85) Somebody has stolen my lipstick.

Frege's first move would be to interpret this as saying that at least one thing satisfies the condition expressed by

(86) x is a person and x has stolen my lipstick.

Thus Frege removed the restriction term, barred all contextual restrictions, and hence trivialised the domain of quantification.

There are two obvious advantages in getting rid of the restriction term: we have fewer separate expressions to deal with, and everything is thrown into the predicate where it can be analysed by way of truth-functors.

However, it is often useful to keep the restriction terms, if only because it makes formulas easier to read. (There are solid technical dividends too, see Feferman [1968b; 1974].) Most logicians who do this follow the advice of Peirce [1885] and use a special style of variable to indicate the restriction. For example set theorists use Greek variables when the restriction is to ordinals. Variables that indicate a special restriction are said to be *sorted* or *sortal*. Two variables marked with the same restriction are said to be *of the same sort*. Logics which use this device are said to be *many-sorted*.

One can also go halfway with Frege and convert the restriction term into another predicate. In this style, 'Every Pole is a Catholic' comes out as a combination of three units: the quantifier word 'Every', the predicate 'x is a Catholic', and a second *relativisation predicate* 'x is a pole'. The mathematical literature is full of *ad hoc* examples of this approach. See for example the bounded quantifiers of number theory in Section 24 below.

When people started to look seriously at other quantifier words besides 'every' and 'some', it became clear that Frege's method of eliminating the restriction term won't always work. For example, the sentence 'Most judges are freemasons' can't be understood as saying that most things satisfy a certain condition. (For a proof of this, and many other examples, see the study of natural language quantifiers by Barwise and Cooper [1981].) For this reason Neil Tennant [Altham and Tennant, 1975] and Barwise [1974] proposed very general formalisms which keep the relativisation predicate separate from the main predicate.

Frege also avoided contextual restrictions. Given his aim, which was to make everything in mathematical reasoning fully explicit, this might seem natural. But it was a bad move. Contextual restrictions do occur, and a logician ought to be prepared to operate with them. In any case various writers have raised philosophical objections to Frege's habit of talking about just everything. Do we really have an undefinable notion of 'object', as Frege supposed? Is it determinate what objects there are? Don't we falsify the meanings of English sentences if we suppose that they state something about everything there is, when on the face of it they are only about Poles?

For a historical study of quantifiers in first-order logic, consult Goldfarb [1979].

10 SATISFACTION

As a convenient and well-known shorthand, we shall say that a person or thing *satisfies* the 1-place predicate ϕ if he or it satisfies the condition which the predicate ϕ expresses. (Notice that we are now allowing the metavariables 'ϕ', 'ψ' etc. to range over predicates as well as sentences and formulas. This shouldn't cause any confusion.)

Many writers put it a little differently. They say that a person or thing satisfies ϕ if the result of putting a name of the person or thing in place of every occurrence of the variable in ϕ is a true sentence. This way of phrasing matters is fine as a first approximation, but it runs into two hazards.

The first hazard is that not everything has a name, even if we allow phrases of the form 'the such-and-such' as names. For example there are uncountably many real numbers and only countably many names.

I can dispose of this objection quickly, as follows. I decree that for purposes of naming arbitrary objects, any ordered pair whose first term is an object and whose second term is the Ayatollah Khalkhali shall be a name of that object. There is a problem about using these names in sentences, but that's just a matter of finding an appropriate convention. So it is clear that if we have an abstract enough notion of what a name is, then every object can have a name.

More conscientious authors have tried to mount reasoned arguments to show that everything is in principle nameable. The results are not always a success. In one paper I recall, the author was apparently under the impression that the nub of the problem was to find a symbol that could be used for naming hitherto nameless objects. After quoting quite a lot of formulas from Quine's *Methods of Logic*, he eventually announced that lower-case italic w can always be used for the purpose. No doubt it can!

There is a second hazard in the 'inserted name' definition of satisfaction. If we allow phrases of the form 'the such-and-such' to count as names, it can happen that on the natural reading, a name means something different within the context of the sentence from what it means in isolation. For example, if my uncle is the mayor of Pinner, and in 1954 he fainted during the opening ceremony of the Pinner Fair, then the mayor of Pinner satisfies the predicate:

(87) In 1954 x fainted during the opening ceremony of the Pinner Fair.

But on the natural reading the sentence

(88) In 1954 the mayor of Pinner fainted during the opening ceremony of the Pinner Fair.

says something quite different and is probably false. One can avoid this phenomenon by sticking to names like 'the present mayor of Pinner' which automatically extract themselves from the scope of surrounding temporal operators (cf. [Kamp, 1971]). But other examples are less easily sorted out. If the programme note says simply 'Peter Warlock wrote this song', then Philip Heseltine, one of whose pen-names was 'Peter Warlock', surely satisfies the predicate

(89) The programme note attributes this song to x.

But my feeling is that on the natural reading, the sentence

(90) The programme note attributes this song to Philip Heseltine

is false. Examples like these should warn us to be careful in applying first-order formalisms to English discourse. (Cf. Bäuerle and Cresswell's chapter 'Propositional Attitudes' to be found in a later Volume of this *Handbook*.)

I turn to some more technical points. We shall need to handle expressions like

(91) x was observed handing a marked envelope to y

which expresses a condition on *pairs* of people or things. It is, I think, quite obvious how to generalize the notion of a 1-place predicate to that of an *n-place predicate*, where n counts the number of distinct individual variables that stand in place of proper names. (Predicates with any positive number of places are also called *open sentences*.) Expression (91) is clearly a 2-place predicate. The only problem is to devise a convention for steering the right objects to the right variables. We do it as follows.

By the *free variables* of a predicate, we mean the individual variables which occur in proper name places in the predicate; so an n-place predicate has n free variables. (In Section 11 we shall have to revise this definition and exclude certain variables from being free.) A predicate with no free variables is called a sentence. We define an *assignment* g to a set of variables (in a situation) to be a function whose domain is that set of variables, with the stipulation that if x is a sorted variable then (in that situation) $g(x)$ meets the restriction which goes with the variable. So for example $g(y_{\text{raccoon}})$ has to be a raccoon.

We say that an assignment g is *suitable for* a predicate ϕ if every free variable of ϕ is in the domain of g. Using the inserted name definition of satisfaction as a temporary expedient, we define: if ϕ is a predicate and g is an assignment which is suitable for ϕ, then g *satisfies* ϕ (in a given situation) iff a true sentence results (in that situation) when we replace each variable x in ϕ by a name of the object $g(x)$.

We shall write

(92) $\alpha/x, \beta/y, \gamma/z, \ldots$

to name the assignment g such that $g(x) = \alpha, g(y) = \beta, g(z) = \gamma$ etc. If $\mathfrak{A}$ is a situation, ϕ a predicate and g an assignment suitable for ϕ, then we write

(93) $\mathfrak{A} \models \phi[g]$

to mean that g satisfies ϕ in the situation $\mathfrak{A}$. The notation (93) is basic for all that follows, so let me give some examples. For simplicity I take $\mathfrak{A}$ to be the real world here and now. The following are true:

(94) $\mathfrak{A} \models$ In the year y, x was appointed Assistant Professor of Mathematics at w at the age of 19 years. [Dr Harvey Friedman/x, 1967/y, Stanford University California/w].

Example (94) asserts that in 1967 Dr Harvey Friedman was appointed Assistant Professor of Mathematics at Stanford University California at the age of 19 years; which must be true because the *Guinness Book of Records* says so.

(95) $\mathfrak{A} \models v$ is the smallest number which can be expressed in two different ways as the sum of two squares. [65/v].

(96) $\mathfrak{A} \models x$ wrote poems about the physical anatomy of x. [Walt Whitman/x].

This notation connects predicates with *objects*, not with names of objects. In (96) it is Mr Whitman himself who satisfies the predicate shown.

In the literature a slightly different and less formal convention is often used. The first time that a predicate ϕ is mentioned, it is referred to, say, as $\phi(y, t)$. This means that ϕ has at most the free variables y and t, and that these variables are to be considered *in that order*. To illustrate, let $\phi(w, x, y)$ be the predicate

(97) In the year y, x was appointed Assistant Professor of Mathematics at w at the age of 19 years.

Then (94) will be written simply as

(98) $\mathfrak{A} \models \phi$ [Stanford University California, Dr Harvey Friedman, 1967].

This handy convention can save us having to mention the variables again after the first time that a predicate is introduced.

There is another variant of (93) which is often used in the study of logics. Suppose that in situation $\mathfrak{A}, g$ is an assignment which is suitable for the predicate ϕ, and S is a sentence which is got from ϕ by replacing each free variable x in ϕ by a name of $g(x)$. Then the truth-value of S is determined by $\mathfrak{A}, g$ and ϕ, and it can be written

(99) $g^*_{\mathfrak{A}}(\phi)$ or $||\phi||_{\mathfrak{A},g}$.

So we have

(100) $\mathfrak{A} \models \phi[g]$ iff $g^*_{\mathfrak{A}}(\phi) = T$.

In (99), $g^*_{\mathfrak{A}}$ can be thought of as a function taking predicates to truth-values. Sometimes it is abbreviated to $g_{\mathfrak{A}}$ or even g, where this leads to no ambiguity.

11 QUANTIFIER NOTATION

Let us use the symbols $x_{\mathrm{boy}}, y_{\mathrm{boy}}$ etc. as sorted variables which are restricted to boys. We shall read the two sentences

(101) $\forall x_{\mathrm{boy}}(x_{\mathrm{boy}}$ has remembered to bring his woggle).

(102) $\exists x_{\mathrm{boy}}(x_{\mathrm{boy}}$ has remembered to bring his woggle).

as meaning exactly the same as (103) and (104) respectively:

(103) Every boy has remembered to bring his woggle.

(104) Some boy has remembered to bring his woggle.

In other words, (101) is true in a situation if and only if in that situation, every member of the domain of quantification of $\forall x_{\mathrm{boy}}$ satisfies the predicate

(105) x_{boy} has remembered to bring his woggle.

Likewise (102) is true if and only if some member of the domain of quantification of $\exists x_{\mathrm{boy}}$ satisfies (105). The situation has to determine what the domain of quantification is, i.e. what boys are being talked about.

The expression $\forall x_{\mathrm{boy}}$ is called a *universal quantifier* and the expression $\exists x_{\mathrm{boy}}$ is called an *existential quantifier*. Because of the restriction 'boy' on the variable, they are called *sorted* or *sortal* quantifiers. The symbols $\forall, \exists$ are called respectively the *universal* and *existential quantifier symbols*; $\forall$ is read 'for all', $\exists$ is read 'for some' or 'there is'.

For unsorted quantifiers using plain variables x, y, z, etc., similar definitions apply, but now the domain of quantification for such a quantifier can be any class of things. Most uses of unsorted quantifiers are so remote from anything in ordinary language that we can't rely on the conventions of speech to locate a domain of quantification for us. So instead we have to assume that *each situation specifies a class which is to serve as the domain of quantification for all unsorted quantifiers.* Then

(106) $\forall x$ (if x is a boy then x has remembered to bring his woggle).

counts as true in a situation if and only if in that situation, every object in the domain of quantification satisfies the predicate

(107) if x is a boy then x has remembered to bring his woggle.

There is a corresponding criterion for the truth of a sentence starting with the unsorted existential quantifier $\exists x$; the reader can easily supply it.

The occurrences of the variable x_{boy} in (101) and (102), and of x in (106), are no longer doing duty for pronouns or marking places where names can be inserted. They are simply part of the quantifier notation. We express this by saying that these occurrences are *bound in* the respective sentences. We also say, for example, that the quantifier at the beginning of (101) *binds* the two occurrences of x_{boy} in that sentence. By contrast an occurrence of a variable in a predicate is called *free in* the predicate if it serves the role we discussed in Sections 9 and 10, of referring to whoever or whatever the predicate expresses a condition on. What we called the *free variables* of a predicate in Section 10 are simply those variables which have free occurrences in the predicate. Note that the concepts 'free' and 'bound' are relative: the occurrence of x_{boy} before 'has' in (101) is bound in (101) but free in (105). Consider also the predicate

(108) x_{boy} forgot his whistle, but $\forall x_{\text{boy}}$ (x_{boy} has remembered to bring his woggle).

Predicate (108) expresses the condition which Billy satisfies if Billy forgot his whistle but every boy has remembered to bring his woggle. So the first occurrence of x_{boy} in (108) is free in (108) but the other two occurrences are bound in (108).

I should recall here the well-known fact that in natural languages, a pronoun can be linked to a quantifier phrase that occurs much earlier, even in a different sentence:

(109) HE: This evening I heard a nightingale in the pear tree.
SHE: It was a thrush—we don't get nightingales here.

In our notation this can't happen. *Our quantifiers bind only variables in themselves and the clause immediately following them.* We express this by saying that the *scope* of an occurrence of a quantifier consists of the quantifier itself and the clause immediately following it; a quantifier occurrence $\forall x$ or $\exists x$ binds all and only occurrences of the same variable x which lie within its scope.

It is worth digressing for a moment to ask why (109) makes life hard for logicians. The crucial question is: just when is the woman's remark 'It was a thrush' a true statement? We want to say that it's true if and only if the object referred to by 'It' is a thrush. But what is there for 'It' to refer to? Arguably the man hasn't referred to any nightingale, he has merely said that there was at least one that he heard in the pear tree. Also we want to say that if her remark is true, then it follows that he heard a thrush in the pear tree. But if this follows, why doesn't it also follow that the nightingale in the pear tree was a thrush? (which is absurd.)

There is a large literature on the problems of cross-reference in natural languages. See for example [Chastain, 1975; Partee, 1978; Evans, 1980]. In the early 1980s Hans Kamp and Irene Heim independently proposed formalisms to handle the matter systematically ([Kamp, 1981; Heim, 1988]; see also [Kamp and Reyle, 1993]). These new formalisms are fundamentally different from first-order logic. Jeroen Groenendijk and Martin Stokhof

[1991] gave an ingenious new semantics for first-order logic which is based on Kamp's ideas and allows a quantifier to pick up a free variable in a later sentence. Their underlying idea is that the meaning of a sentence is the change which it makes to the information provided by earlier sentences in the conversation. This opens up new possibilities, but it heads in a very different direction from the usual first-order logic.

Returning to first-order logic, consider the sentence

(110) $\exists x_{\text{boy}}(x_{\text{boy}}$ kissed Brenda).

This sentence can be turned into a predicate by putting a variable in place of 'Brenda'. Naturally the variable we use has to be different from x_{boy}, or else it would get bound by the quantifier at the beginning. Apart from that constraint, any variable will do. For instance:

(111) $\exists x_{\text{boy}}(x_{\text{boy}}$ kissed $y_{\text{girlwithpigtails}})$.

We need to describe the conditions in which Brenda satisfies (111). Brenda must of course be a girl with pigtails. She satisfies (111) if and only if there is a boy β such that the assignment

(112) β/x_{boy}, Brenda/$y_{\text{girlwithpigtails}}$

satisfies the predicate 'x_{boy} kissed $y_{\text{girlwithpigtails}}$'. Formal details will follow in Section 14 below.

12 AMBIGUOUS CONSTANTS

In his *Wissenschaftslehre II* [1837, Section 147] Bernard Bolzano noted that we use demonstrative pronouns at different times and places to refer now to this, now to that. He continued:

> Since we do this anyhow, it is worth the effort to undertake this procedure with full consciousness and with the intention of gaining more precise knowledge about the nature of such propositions by observing their behaviour with respect to truth. Given a proposition, we could merely inquire whether it is true or false. But some very remarkable properties of propositions can be discovered if, in addition, we consider the truth values of all those propositions which can be generated from it, if we take some of its constituent ideas as variable and replace them by any other ideas whatever.

We can abandon to the nineteenth century the notion of 'variable ideas'. What Bolzano did in fact was to introduce *totally ambiguous symbols.* When a writer uses such a symbol, he has to indicate what it means, just as he has to make clear what his demonstrative pronouns refer to. In our terminology, the situation must fix the meanings of such symbols. Each totally ambiguous symbol has a certain grammatical type, and the meaning supplied must fit the grammatical type; but that apart, anything goes.

Let us refer to a sentence which contains totally ambiguous symbols as a *sentence schema*. Then an *argument schema* will consist of a string of sentence schemas called *premises*, followed by the word '*Therefore*', followed by a sentence schema called the *conclusion*. A typical argument schema might be:

(113) a is more X than b. b is more X than c. *Therefore* a is more X than c.

A traditional logician would have said that (113) is a valid argument schema if and only if all its instances are valid arguments (cf. (1) in the Introduction above). Bolzano said something different. Following him, we shall say that (113) is *Bolzano-valid* if for every situation in which a, b, c are interpreted as names and X is interpreted as an adjective, either one or more of the premises are not true, or the conclusion is true. We say that the premises in (113) *Bolzano-entail* the conclusion if (113) is Bolzano-valid.

Note the differences. For the traditional logician entailment is from sentences to sentences, not from sentence schemas to sentence schemas. Bolzano's entailment is between schemas, not sentences, and moreover he defines it without mentioning entailment between sentences. The schemas become sentences of a sort when their symbols are interpreted, but Bolzano never asks whether these sentences "can't be true without certain other sentences being true" (to recall our definition of entailment in the Introduction)—he merely asks when they *are* true.

The crucial relationship between Bolzano's ideas and the traditional ones is that *every instance of a Bolzano-valid argument schema is a valid argument*. If an argument is an instance of a Bolzano-valid argument schema, then that fact itself is a reason why the premises can't be true without the conclusion also being true, and so the argument is valid. The traditional logician may want to add a caution here: the argument need not be *logically* valid unless the schema is Bolzano-valid for *logical* reasons—whatever we take 'logical' to mean. Tarski [1936] made this point. (Let me take the opportunity to add that recent discussions of the nature of logical consequence have been clouded by some very unhistorical readings of [Tarski, 1936]. Fortunately there is an excellent historical analysis by Gómez-Torrente [1996].)

In first-order logic we follow Bolzano and study entailments between schemas. We use two kinds of totally ambiguous constants. The first kind are the *individual constants*, which are normally chosen from lower-case letters near the beginning of the alphabet: a, b, c etc. These behave grammatically as singular proper names, and are taken to stand for objects. The other kind are the *predicate* (or *relation*) *constants*. These are usually chosen from the letters P, Q, R etc. They behave as verbs or predicates, in the following way. To specify a meaning for the predicate constant P, we could write

(114) $Pxyz$ means x aimed at y and hit z.

The choice of variables here is quite arbitrary, so (114) says the same as:

(115) $Pyst$ means y aimed at s and hit t.

We shall say that under the interpretation (114), an ordered 3-tuple $\langle\alpha, \beta, \gamma\rangle$ of objects *satisfies* P if and only if the assignment

(116) $\alpha/x, \beta/y, \gamma/z$

satisfies the predicate 'x aimed at y and hit z'. So for example the ordered 3-tuple ⟨Bert, Angelo, Chen⟩ satisfies P under the interpretation (114) or (115) if and only if Bert aimed at Angelo and hit Chen. (We take P to be satisfied by ordered 3-tuples rather than by assignments because, unlike a predicate, the symbol P comes without benefit of variables.) The collection of all ordered 3-tuples which satisfy P in a situation where P has the interpretation (114) is called the *extension* of P in that situation. In general a collection of ordered n-tuples is called an *n-place relation*.

Since P is followed by three variables in (114), we say that P in (114) is serving as a *3-place predicate constant*. One can have n-place predicate constants for any positive integer n; the extension of such a constant in a situation is always an n-place relation. In theory a predicate constant could be used both as a 3-place and as a 5-place predicate constant in the same setting without causing mishap, but in practice logicians try to avoid doing this.

Now consider the sentence

(117) $\forall x$ (if Rxc then x is red).

with 2-place predicate constant R and individual constant c. What do we need to be told about a situation $\mathfrak{A}$ in order to determine whether (117) is true or false in $\mathfrak{A}$? The relevant items in $\mathfrak{A}$ seem to be:

(a) the domain of quantification for $\forall x$.

(b) the object named by the constant c. (Note: it is irrelevant what meaning c has over and above naming this object, because R will be interpreted by a predicate.) We call this object $I_{\mathfrak{A}}(c)$.

(c) the extension of the constant R. (Note: it is irrelevant what predicate is used to give R this extension; the extension contains all relevant information.) We call this extension $I_{\mathfrak{A}}(R)$.

(d) the class of red things.

In Section 14 we shall define the important notion of a *structure* by extracting what is essential from (a)–(d). Logicians normally put into the definition of 'structure' some requirements that are designed to make them simpler to handle. Before matters get buried under symbolism, let me say what these requirements amount to in terms of $\mathfrak{A}$. (See Appendix C below for the set-theoretic notions used.)

1. There is to be a collection of objects called the *domain* of $\mathfrak{A}$, in symbols $|\mathfrak{A}|$.

2. $|\mathfrak{A}|$ is the domain of quantification for all unsorted quantifiers. Two sorted quantifiers with variables of the same sort (if there are any) always have the same domain of quantification, which is included in $|\mathfrak{A}|$.

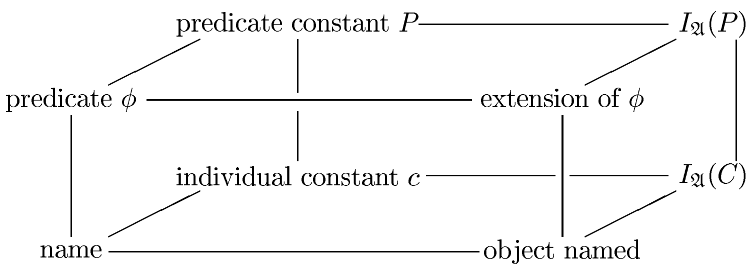
3. For every individual constant c, the interpretation $I_{\mathfrak{A}}(c)$ is a member of $|\mathfrak{A}|$; for every predicate constant R, the relation $I_{\mathfrak{A}}(R)$ is a relation on $|\mathfrak{A}|$.

4. Some authors require $|\mathfrak{A}|$ to be a pure set. Most authors require it to have at least one member. A very few authors (e.g. [Carnap, 1956; Hintikka, 1955]) require it to be at most countable.

Requirements (1)–(3) mean in effect that first-order logicians abandon any pretence of following the way that domains of quantification are fixed in natural languages. Frege's device of Section 9 (e.g. (84)) shows how we can meet these requirements and still say what we wanted to say, though at greater length. Requirements (4) are an odd bunch; I shall study their reasons and justifications in due course below.

Logicians also allow one important relaxation of (1)–(4). They permit an n-place predicate symbol to be interpreted by *any* n-place relation on the domain, not just one that comes from a predicate. Likewise they permit an individual constant to stand for any member of the domain, regardless of whether we can identify that member. The point is that the question whether *we* can describe the extension or the member is totally irrelevant to the question what is true in the structure.

Note here the 3-way analogy



The front face of this cube is essentially due to Frege. Would he have accepted the back?

No, he would not. In 1899 Hilbert published a study of the axioms of geometry. Among other things, he asked questions of the form 'Do axioms A, B, C together entail axiom D?' (The famous problem of the independence of Euclid's parallel postulate is a question of this sort.) Hilbert answered these questions by regarding the axioms as schemas containing ambiguous signs, and then giving number-theoretic interpretations which made the premises A, B and C true but the conclusion D false. Frege read the book [Hilbert, 1899] and reacted angrily. After a brief correspondence with Hilbert (Frege and Hilbert [1899–1900], he published a detailed critique [1906], declaring [Frege, 1971, p. 66]: "Indeed, if it were a matter of deceiving oneself and others, there would be no better means than ambiguous signs."

Part of Frege's complaint was that Hilbert had merely shown that certain argument schemas were not Bolzano-valid; he had not shown that axioms A, B and C, taken literally as statements about points, lines etc. in real space, do not entail axiom D taken literally. This is true and need not detain us—Hilbert had answered the questions he wanted to answer. Much more seriously, Frege asserted that Hilbert's propositions, being ambiguous, did not express determinate thoughts and hence could not serve as the premises or conclusions of inferences. In short, Frege refused to consider Bolzano-valid argument schemas as any kind of valid argument. So adamant was he about this that he undertook to translate the core of Hilbert's reasoning into what he considered an acceptable form which never mentioned schematic sentences. This is not difficult to do— it is a matter of replacing statements of the form 'Axiom A entails axiom B' by statements of the form 'For all relations P and R, if P and R do this then they do that'. But the resulting translation is quite unreadable, so good mathematics is thrown away and all for no purpose.

Frege's rejection of ambiguous symbols is part and parcel of his refusal to handle indexical expressions; see [Perry, 1977] for some discussion of the issue. It is sad to learn that the grand architect of modern logic fiercely rejected the one last advance which was needed to make his ideas fruitful.

In fact it took some years for logicians to accept the use of ambiguous symbols in the semantics of first-order logic. For example Tarski's paper [1936] on logical deduction made no use of them; Tarski found another device with the same effect (at the cost of adapting the word 'model' to mean '*re*-interpretation' rather than 'interpretation'). But in his model-theoretic work of the 1950s and later, Tarski used ambiguous constants wholesale in the modern fashion, as a form of indexical. (Cf. [Hodges, 1985/86].)

13 FIRST-ORDER SYNTAX FORMALISED

The main purpose of this section and the next is to extract the formal content of Sections 9–12 above. I give the definitions first under the assumption that there are no sorted variables. Also I ignore for the moment the fact that some first-order logicians use = and function symbols. Section 18 below will be more broad-minded.

A *similarity type* is defined to be a set of individual constants together with a set of predicate constants; each predicate constant is assumed to be labelled somehow to indicate that it is an n-place predicate constant, for some positive integer n. Some writers include the n as a superscript: R^{133} is a 133-place predicate constant.

We shall define the *first-order language* L of *similarity type* X. For definiteness, L shall be an ordered triple $\langle X, T(X), F(X)\rangle$ where X is the similarity type, and $T(X)$ and $F(X)$ are respectively the set of all terms and formulas of similarity type X (known more briefly as the terms and formulas of L). Grammatically speaking, the terms of L are its noun phrases and the formulas are its sentences. Metavariables σ, τ will range over terms, and metavariables ϕ, ψ, χ will range over formulas.

We start the definition by defining the *variables* to be the countably many symbols

(118) $x_0, x_1, x_2, \ldots.$

Unofficially everybody uses the symbol x, y, z etc. as variables. But in the spirit of Section 4 above, these can be understood as metavariables ranging over variables. The *terms* of L are defined to be the variables of L and the individual constants in X.

An *atomic formula* of L is an expression of form $P(\sigma_1, \ldots, \sigma_n)$ where P is an n-place predicate constant in X and $\sigma_1, \ldots, \sigma_n$ are terms of L. The class of *formulas* of L is defined inductively, and as the induction proceeds we shall define also the set of subformulas of the formula ϕ, and the set $FV(\phi)$ of free variables of ϕ:

(a) Every atomic formula ϕ of L is a formula of L; it is its only subformula, and $FV(\phi)$ is the set of all variables which occur in ϕ. $\bot$ is a formula of L; it is its only subformula, and $FV(\bot)$ is empty.

(b) Suppose ϕ and ψ are formulas of L and x is a variable. Then: $\neg\phi$ is a formula of L; its subformulas are itself and the subformulas of ϕ; $FV(\neg\phi)$ is $FV(\phi)$. Also $(\phi \wedge \psi), (\phi \vee \psi), (\phi \to \psi)$ and $(\phi \leftrightarrow \psi)$ are formulas of L; the subformulas of each of these formulas are itself, the subformulas of ϕ and the subformulas of ψ; its free variables are those of ϕ together with those of ψ. Also $\forall x\phi$ and $\exists x\phi$ are formulas of L; for each of these, its subformulas are itself and the subformulas of ϕ; its free variables are those of ϕ excluding x.

(c) Nothing is a formula of L except as required by (a) and (b).

The *complexity* of a formula ϕ is defined to be the number of subformulas of ϕ. This definition disagrees with that in Section 3, but it retains the crucial property that every formula has a higher complexity than any of its proper subformulas. (The *proper subformulas* of ϕ are all the subformulas of ϕ except ϕ itself.) A formula is said to be *closed*, or to be a *sentence*, if it has no free variables. Closed formulas correspond to sentences of English, non-closed formulas to predicates or open sentences of English. Formulas of a formal language are sometimes called *well-formed formulas* or *wffs* for short.

If ϕ is a formula, x is a variable and τ is a term, then there is a formula $\phi[\tau/x]$ which 'says the same thing about the object τ as ϕ says about the object x'. At a first approximation, $\phi[\tau/x]$ can be described as the formula which results if we put τ in place of each free occurrence of x in ϕ; when this description works, we say τ is *free for x in ϕ* or *substitutable* for x in ϕ. Here is an example where the approximation doesn't work: ϕ is $\exists yR(x,y)$ and τ is y. If we put y for x in ϕ, the resulting formula $\exists yR(y,y)$ says nothing at all about 'the object y', because the inserted y becomes bound by the quantifier $\exists y$—a phenomenon known as *clash of variables*. In such cases we have to define $\phi[\tau/x]$ to be $\exists zR(y,z)$ where z is some other variable. (There is a good account of this messy matter in Bell and Machover [1977, Chapter 2, Section 3].)

Note the useful shorthand: if ϕ is described at its first occurrence as $\phi(x)$, then $\phi(\tau)$ means $\phi[\tau/x]$. Likewise if ϕ is introduced as $\phi(y_1,\ldots,y_n)$ then $\phi(\tau_1,\ldots,\tau_n)$ means the formula which says about the objects $\tau_1,\ldots,\tau_n$ the same thing as ϕ says about the objects $y_1,\ldots,y_n$.

Not much in the definitions above needs to be changed if you want a system with sorted variables. You must start by deciding what kind of sortal system you want. There will be a set S of sorts s,t etc., and for each sort s there will be sorted variables x_0^s, s_1^s, x_2^s etc. But then (a) do you want every object to belong to some sort? If so, the similarity type must assign each individual constant to at least one sort. (b) Do you want the sorts to be mutually exclusive? Then the similarity type must assign each individual constant to at most one sort. (c) Do you want to be able to say 'everything', rather than just 'everything of such-and-such a sort'? If not then the unsorted variables (118) should be struck out.

Some formal languages allow restricted quantification. For example in languages designed for talking about numbers, we have formulas $(\forall x < y)\phi$ and $(\exists x < y)\phi$, read respectively as 'For all numbers x less than y, ϕ' and 'There is a number x less than y such that ϕ'. These expressions can be regarded as metalanguage abbreviations for $\forall x(x < y \to \phi)$ and $\exists x(x < y \wedge \phi)$ respectively (where '$x < y$' in turn is an abbreviation for '$< (x, y)$'). Or we can alter the definition of 'formula of L' to allow restricted quantifiers in L itself.

One often sees abbreviations such as '$\forall xy\phi$' or '$\exists \vec{z}\phi$'. These are metalanguage abbreviations. $\forall xy$ is short for $\forall x \forall y$. $\vec{z}$ means a finite sequence $z_1, \ldots,$ z_n. Furthermore, the abbreviations of Section 4 remain in force.

All the syntactic notions described in this section can be defined using only concrete instances of the induction axiom as in Section 3 above.

14 FIRST-ORDER SEMANTICS FORMALISED

We turn to the definition of structures. (They are also known as *models*—but it is better to reserve this term for the context 'model of ϕ'.) Let L be a language with similarity type X. Then an L-*structure* $\mathfrak{A}$ is defined to be an ordered pair $\langle A, I \rangle$ where:

1. A is a class called the *domain* of $\mathfrak{A}$, in symbols $|\mathfrak{A}|$. The elements of A are called the *elements* of $\mathfrak{A}$, and the cardinality of A is called the *cardinality* of $\mathfrak{A}$. So for example we call $\mathfrak{A}$ *finite* or *empty* if A is finite or empty. Many writers use the convention that A, B and C are the domains of $\mathfrak{A}, \mathfrak{B}$ and $\mathfrak{C}$ respectively.

2. I is a function which assigns to each individual constant c of X an element $I(c)$ of A, and to each n-place predicate symbol R of X an n-place relation $I(R)$ on A. I is referred to as $I_{\mathfrak{A}}$.

Structure means: L-structure for some language L.

If Z is a set of variables, then an *assignment* to Z in $\mathfrak{A}$ is defined to be a function from Z to A. If g is an assignment to Z in $\mathfrak{A}, x$ is a variable not in Z and α is an element of $\mathfrak{A}$, then we write

(119) $g, \alpha/x$

for the assignment h got from g by adding x to g's domain and putting $h(x) = \alpha$. (Some writers call assignments *valuations*.)

For each assignment g in $\mathfrak{A}$ and each individual constant c we define $c[g]$ to be the element $I_{\mathfrak{A}}(c)$. For each variable x and assignment g whose domain contains x, we define $x[g]$ to be the element $g(x)$. Then $\tau[g]$ is 'the element named by the term τ under the assignment g'.

For each formula ϕ of L and each assignment g to the free variables of ϕ in $\mathfrak{A}$, we shall now define the conditions under which $\mathfrak{A} \models \phi[g]$ (cf. (93) above). The definition is by induction on the complexity of ϕ.

(a) If R is an n-place predicate constant in X and $\tau_1, \ldots, \tau_n$ are terms, then $\mathfrak{A} \models R(\tau_1, \ldots, \tau_n)$ iff the ordered n-tuple $\langle \tau_1[g], \ldots, \tau_n[g] \rangle$ is in $I_{\mathfrak{A}}(R)$.

(b) It is never true that $\mathfrak{A} \models \bot$.

(c) $\mathfrak{A} \models \neg\phi[g]$ iff it is not true that $\mathfrak{A} \models \phi[g]$.
$\mathfrak{A} \models \phi \wedge \psi[g]$ iff $\mathfrak{A} \models \phi[g_1]$ and $\mathfrak{A} \models \psi[g_2]$, where g_1 and g_2 are the results of restricting g to the free variables of ϕ and ψ respectively.
Etc. as in (23).

(d) If x is a free variable of ϕ, then:
$\mathfrak{A} \models \forall x\phi[g]$ iff for every element α of $A, \mathfrak{A} \models \phi[g, \alpha/x]$;
$\mathfrak{A} \models \exists x\phi[g]$ iff for at least one element α of $A, \mathfrak{A} \models \phi[g, \alpha/x]$.
If x is not a free variable of ϕ, then $\mathfrak{A} \models \forall x\phi[g]$ iff $\mathfrak{A} \models \phi[g]$, and $\mathfrak{A} \models \exists x\phi[g]$ iff $\mathfrak{A} \models \phi[g]$.

We say an assignment g in $\mathfrak{A}$ is *suitable for* the formula ϕ if every free variable of ϕ is in the domain of g. If g is suitable for ϕ, we say that $\mathfrak{A} \models \phi[g]$ if and only if $\mathfrak{A} \models \phi[h]$, where h comes from g by throwing out of the domain of g those variables which are not free variables of ϕ.

If ϕ is a sentence, then ϕ has no free variables and we can write just $\mathfrak{A} \models \phi$ in place of $\mathfrak{A} \models \phi[\ ]$. This notation agrees with (22) above. When $\mathfrak{A} \models \phi$, we say that $\mathfrak{A}$ is a *model of* ϕ, or that ϕ is *true in* $\mathfrak{A}$. '$\mathfrak{A} \models \phi[g]$' can be pronounced '*g satisfies* ϕ in $\mathfrak{A}$'.

To anybody who has mastered the symbolism it should be obvious that clauses (a)–(d) really do determine whether or not $\mathfrak{A} \models \phi$, for every L-structure $\mathfrak{A}$ and every sentence ϕ of L. *If* $\mathfrak{A}$ *is a set* then we can formalise the definition in the language of set theory and prove that it determines $\models$ uniquely, using only quite weak set-theoretic axioms (cf. [Barwise, 1975, Chapter 3]). Set structures are adequate for most applications of first-order logic in mathematics, so that many textbooks simply state without apology that a structure has to be a set. We shall return to this point in Section 17 below.

The definition of $\models$ given above is called the *truth-definition*, because it specifies exactly when a symbolic formula is to count as 'true in' a structure. It solves no substantive problems about what is true—we are just as much in the dark about the Riemann hypothesis or the Reichstag fire after writing it down as we were before. But it has attracted a lot of attention as a possible answer to the question of what is Truth. Many variants of it have appeared in the literature, which can cause anguish to people anxious to get to the heart of the matter. Let me briefly describe three of these variants; they are all mathematically equivalent to the version given above. (Cf. Leblanc [Volume 2 of this *Handbook*].)

In the first variant, *assignments are sequences*. More precisely an assignment in $\mathfrak{A}$ is defined to be a function g from the natural numbers N to the domain A of $\mathfrak{A}$. Such a function can be thought of as an infinite sequence $\langle g(0), g(1), g(2), \ldots \rangle$. The element $g(i)$ is assigned to the ith variable x_i, so that $x_i[g]$ is defined to be $g(i)$. In (c) and (d) we have to make some changes for the purely technical reason that g assigns elements to *every* variable and not just those free in ϕ. In (c) the clause for $\phi \wedge \psi$ becomes

$$\mathfrak{A} \vDash \phi \wedge \psi[g] \quad \text{iff} \quad \mathfrak{A} \vDash \phi[g] \text{ and } \mathfrak{A} \vDash \psi[g],$$

which is an improvement (and similarly with $(\phi \vee \psi), (\phi \to \psi)$ and $(\phi \leftrightarrow \psi)$). But (d) becomes distorted, because g already makes an assignment to the quantified variable x; this assignment is irrelevant to the truth of $\mathfrak{A} \vDash \forall x \phi[g]$, so we have to discard it as follows. For each number i and element α of $\mathfrak{A}$, let $g(\alpha/i)$ be the assignment h which is exactly like g except that $h(i) = \alpha$. Then (d) is replaced by:

(d′) For each variable x_i : $\mathfrak{A} \vDash \forall x_i \phi[g]$ iff for every element α of A, $\mathfrak{A} \vDash \phi[g(\alpha/i)]$.

together with a similar clause for $\exists x_i \phi$.

In the second variant, we copy (24) and define the *truth-value* of ϕ in $\mathfrak{A}$, $||\phi||_{\mathfrak{A}}$, to be the set of all assignments g to the free variables of ϕ such that $\mathfrak{A} \vDash \phi[g]$. When ϕ is a sentence, there is only one assignment to the free variables of ϕ, namely the empty function 0; so $||\phi||_{\mathfrak{A}}$ is $\{0\}$ if ϕ is true in $\mathfrak{A}$, and the empty set (again 0) if ϕ is false in $\mathfrak{A}$. This variant is barely more than a change of notation. Instead of '$\mathfrak{A} \vDash \phi[g]$' we write '$g \in ||\phi||_{\mathfrak{A}}$'. The clauses (a)–(d) can be translated easily into the new notation.

Some writers combine our first and second variants, taking $||\phi||_{\mathfrak{A}}$ to be the set of all sequences g such that $\mathfrak{A} \vDash \phi[g]$. In this style, the clause for $\phi \wedge \psi$ in (c) becomes rather elegant:

$$||\phi \wedge \psi||_{\mathfrak{A}} = ||\phi||_{\mathfrak{A}} \cap ||\psi||_{\mathfrak{A}}.$$

However, when ϕ is a sentence the definition of 'ϕ is true in $\mathfrak{A}$' becomes 'every sequence is in $||\phi||_{\mathfrak{A}}$', or equivalently 'at least one sequence is in $||\phi||_{\mathfrak{A}}$'. I have heard students repeat this definition with baffled awe as if they learned it in the Eleusinian Mysteries.

The third variant dispenses with assignments altogether and adds new constant names to the language L. Write L(c) for the language got from L by adding c as an extra individual constant. If $\mathfrak{A}$ is an L-structure and α is an element of $\mathfrak{A}$, write $(\mathfrak{A}, \alpha)$ for the L(c)-structure $\mathfrak{B}$ which is the same as $\mathfrak{A}$ except that $I_{\mathfrak{B}}(c) = \alpha$. If ϕ is a formula of L with just the free variable x, one can prove by induction on the complexity of ϕ that

(120) $(\mathfrak{A}, \alpha) \vDash \phi[c/x]$ iff $\mathfrak{A} \vDash \phi[\alpha/x]$.

(Warning: $[c/x]$ on the left is a substitution in the formula ϕ; α/x on the right is an assignment to the variable x.) The two sides in (120) are just different ways of expressing that α satisfies ϕ in $\mathfrak{A}$. Hence we have

(121) $\mathfrak{A} \vDash \forall x\phi$ iff for every element α of $\mathfrak{A}, (\mathfrak{A}, \alpha) \vDash \phi[c/x]$,

and a similar clause for $\exists x\phi$. In our third variant, (121) is taken as the *definition* of $\vDash$ for sentences of form $\forall x\phi$. This trick sidesteps assignments. Its disadvantage is that we have to alter the language and the structure each time we come to apply clause (d). The great merit of assignments is that they enable us to keep the structure fixed while we wiggle around elements in order to handle the quantifiers.

There are L-structures whose elements are all named by individual constants of L. For example, the natural numbers are sometimes understood as a structure in which every number n is named by a numeral constant $\overline{n}$ of the language. For such structures, *and only for such structures*, (121) can be replaced by

(122) $\mathfrak{A} \vDash \forall x\phi$ iff for every individual constant c of $\mathrm{L}, \mathfrak{A} \vDash \phi[c/x]$.

Some writers confine themselves to structures for which (122) applies.

Alfred Tarski's famous paper on the concept of truth in formalised languages [1935] was the first paper to present anything like our definition of $\vDash$. Readers should be aware of one vital difference between his notion and ours. His languages have no ambiguous constants. True, Tarski says they have constants. But he explains that by 'constants' he means negation signs, quantifier symbols and suchlike, together with symbols of fixed meaning such as the inclusion sign $\subseteq$ in set theory. (See Section 20 below on symbols with an 'intended interpretation'.) The only concession that Tarski makes to the notion of an L-structure is that he allows the domain of elements to be any class, not necessarily the class of everything. Even then he says that relativising to a particular class is 'not essential for the understanding of the main theme of this work'! (Cf. pages 199, 212 of the English translation of [Tarski, 1935].) Carnap's truth-definition [1935] is also little sideways from modern versions.

There is no problem about adapting Tarski's definition to our setting. It can be done in several ways. Probably the simplest is to allow some of his constants to turn ambiguous; then his definition becomes our first variant.

Finally I should mention structures for many-sorted languages, if only to say that no new issues of principle arise. If the language L has a set S of sorts, then for each sort s in S, an L-structure $\mathfrak{A}$ must carry a class $s(\mathfrak{A})$ of *elements of sort* s. In accordance with Section 12, $s(\mathfrak{A})$ must be included in $|\mathfrak{A}|$. If the individual constant c is of sort s, then $I_{\mathfrak{A}}(c)$ must be an element of $s(\mathfrak{A})$. If we have required that every element should be of at least one sort, then $|\mathfrak{A}|$ must be the union of the classes $s(\mathfrak{A})$.

15 FIRST-ORDER IMPLICATIONS

Let me make a leap that will seem absurd to the Traditional Logician, and define sequents with infinitely many premises.

Suppose L is a first-order language. By a *theory in* L we shall mean a set of sentences of L—it can be finite or infinite. The metavariables $\Delta, \Gamma, \Theta, \Lambda$ will range over theories. If Δ is a theory in L and $\mathfrak{A}$ is an L-structure, we say that $\mathfrak{A}$ is a *model of* Δ if $\mathfrak{A}$ is a model of every sentence in Δ.

For any theory Δ in L and sentence ϕ of L, we define

(123) $\Delta \vDash \phi$ ('Δ *logically implies* ϕ', 'ϕ is a *logical consequence* of Δ')

to mean that every L-structure which is a model of Δ is also a model of ϕ. If Δ has no models, (123) is reckoned to be true by default. A *counterexample* to (123) is an L-structure which is a model of Δ but not of ϕ. We write

(124) $\vDash \phi$ ('ϕ is *logically valid*')

to mean that every L-structure is a model of ϕ; a *counterexample* to (124) is an L-structure which is not a model of ϕ. The expressions (123) and (124) are called *sequents*. This definition of logical implication was first set down by Tarski [1936], though it only makes precise what Bolzano [1837, Section 155] and Hilbert [1899] already understood.

Warning: (123) is a definition of logical consequence *for first-order schemas*. It doesn't make sense as a definition of logical consequence between meaningful sentences, even when the sentences are written in first-order notation; logical consequence might hold between the sentences for reasons not expressed in the first-order notation. This is obvious: let 'p' stand for your favourite logical truth, and consider '$\models p$'. I mention this because I have seen a small river of philosophical papers which criticise (123) under the impression that it is intended as a definition of logical consequence between sentences of English (they call it the 'model-theoretic definition of logical consequence'). In one case where I collared the author and traced the mistake to source, it turned out to be a straight misreading of that excellent textbook [Enderton, 1972]; though I am not sure the author accepted my correction. One can track down some of these confusions to the terminology of Etchemendy [1990], who uses phrases such as 'the set of logical truths of any given first-order language' [Etchemendy, 1990, p. 148] to mean those sentences of a *fully interpreted* first-order language which are (in Etchemendy's sense) intuitively logically true. In his Chapter 11 especially, Etchemendy's terminology is way out of line with that of the authors he is commenting on.

If the language L has at least one individual constant c, then every L-structure must have an element $I_{\mathfrak{A}}(c)$, so the domain of $\mathfrak{A}$ can't be empty. It follows that in this language the sentence $\exists x \neg \bot$ must be logically valid, so we can 'prove' that at least one thing exists.

On the other hand if L has no individual constants, then there is an L-structure whose domain is empty. This is not just a quirk of our conventions: one can quite easily think of English sentences uttered in contexts where the natural domain of quantification happens to be empty. In such a language L, $\exists x \neg \bot$ is not logically valid.

This odd state of affairs deserves some analysis. Suppose L does have an individual constant c. By the Bolzano–Tarski definition (123), when we consider logical implication in L we are only concerned with structures in which c names something. In other words, the Bolzano–Tarski definition slips into every argument a tacit premise that *every name does in fact name something*. If we wanted to, we could adapt the Traditional Logician's notion of a valid argument in just the same way. For a traditional example, consider

(125) Every man runs. *Therefore* Socrates, if he is a man, runs.

On the traditional view, (125) is not a valid argument—it could happen that every man runs and yet there is no such entity as Socrates. On the Bolzano–Tarski view we must consider only situations in which 'Socrates' names something or someone, and on that reckoning, (125) is valid. (According to Walter Burleigh in the fourteenth century, (125) is not valid outright, but it is valid at the times when Socrates exists. Cf. Bocheński [1970, p. 193]; I have slightly altered Burleigh's example. I don't know how one and the same argument can be valid at 4 p.m. and invalid at 5 p.m.).

Once this much is clear, we can decide whether we want to do anything about it. From the Traditional Logician's point of view it might seem sensible to amend the Bolzano–Tarski definition. This is the direction which *free logic* has taken. Cf. Bencivenga, (Volume 7 of this *Handbook*).

The mainstream has gone the other way. Non-referring constants are anathema in most mathematics. Besides, Hilbert-style calculi with identity always have $\exists x(x = x)$ as a provable formula. (See Remark 6 in Appendix A below. On the other hand semantic tableau systems which allow empty structures, such as Hodges [1977], are arguably a little simpler and more natural than versions which exclude them.) If $\exists x \neg \bot$ is logically valid in some languages and not in others, the easiest remedy is to make it logically valid in all languages, and we can do that by *requiring all structures to have non-empty domains*. Henceforth we shall do so (after pausing to note that Schröder [1895, p. 5] required all structures to have at least two elements).

Let us review some properties of $\models$. Analogues of Theorems 1–4 (allowing infinitely many premises!) and Theorem 5 of Section 5 now hold. The relevant notion of logical equivalence is this: the formula ϕ is *logically equivalent* to the formula ϕ if for every structure $\mathfrak{A}$ and every assignment g in $\mathfrak{A}$ which is suitable for both ϕ and ψ, $\mathfrak{A} \models \phi[g]$ if and only if $\mathfrak{A} \models \psi[g]$. For example

(126) $\forall x\phi$ is logically equivalent to $\neg\exists x\neg\phi$,
$\exists x\phi$ is logically equivalent to $\neg\forall x\neg\phi$.

A formula is said to be *basic* if it is either atomic or the negation of an atomic formula. A formula is in *disjunctive normal form* if it is either $\perp$ or a disjunction of conjunctions of basic formulas. One can show:

(127) *Every formula of* L *is logically equivalent to a formula of* L *with the same free variables, in which all quantifiers are at the left-hand end, and the part after the quantifiers is in disjunctive normal form.*

A formula with its quantifiers all at the front is said to be in *prenex form.* (In Section 25 below we meet Skolem normal forms, which are different from (127) but also prenex.)

Proof calculi for propositional logic are generally quite easy to adapt to predicate logic. Sundholm (Volume 2 of this *Handbook*) surveys the possibilities. Usually in predicate logic one allows arbitrary formulas to occur in a proof, not just sentences, and this can make it a little tricky to say exactly what is the informal idea expressed by a proof. (This applies particularly to Hilbert-style calculi; cf. Remarks 4 and 5 in Appendix A below. Some calculi paper over the difficulty by writing the free variables as constants.) When one speaks of a formal calculus for predicate logic as being *sound* or *complete* (cf. Section 7 above), one always ignores formulas which have free variables.

Gentzen's natural deduction calculus can be adapted to predicate logic simply by adding four rules, namely introduction and elimination rules for $\forall$ and $\exists$. The *introduction rule* for $\exists$ says:

(128) From $\phi[\tau/x]$ infer $\exists x\phi$.

(If the object τ satisfies ϕ, then at least one thing satisfies ϕ.) The *elimination rule* for $\exists$ says:

(129) Given a proof of ψ from $\phi[y/x]$ and assumptions $\chi_1, \ldots, \chi_n$, where y is not free in any of $\exists x\phi, \psi, \chi_1, \ldots, \chi_n$, deduce ψ from $\exists x\phi$ and $\chi_1, \ldots, \chi_n$.

The justification of (129) is of some philosophical interest, as the following example will show. We want to deduce an absurdity from the assumption that there is a greatest integer. So we let y be a greatest integer, we get a contradiction $y < y + 1 \leq y$, whence $\perp$. Then by (129) we deduce $\perp$ from $\exists x$ (x is a greatest integer). Now the problem is: How can we possibly 'let y be a greatest integer', since there aren't any? Some logicians exhort us to '*imagine* that y is a greatest integer', but I always found that this one defeats my powers of imagination.

The Bolzano–Tarski definition of logical implication is a real help here, because it steers us away from matters of 'If it were the case that ...' towards questions about what actually is the case in structures which do exist. We have to decide how natural deduction proofs are supposed to match the Bolzano–Tarski definition, bearing in mind that formulas with free variables may occur. The following interpretation is the right one: the existence of a natural deduction proof with conclusion ψ and premises $\chi_1, \ldots, \chi_n$ should tell us that for every structure $\mathfrak{A}$ and every assignment g in $\mathfrak{A}$ which is suitable for all of $\psi, \chi_1, \ldots, \chi_n$, we have $\mathfrak{A} \vDash (\chi_1 \wedge \cdots \wedge \chi_n \to \psi)[g]$. (This is *not* obvious— for Hilbert-style calculi one has to supply a quite different rationale, cf. Remark 5 on Hilbert-style calculi in Appendix A.)

Now we can justify (129). Let $\mathfrak{A}$ be a structure and g an assignment in $\mathfrak{A}$ which is suitable for $\exists x\phi, \chi_1, \ldots, \chi_n$ and ψ. We wish to show that:

(130) $\mathfrak{A} \vDash (\exists x\phi \wedge \chi_1 \wedge \cdots \wedge \chi_n \to \psi)[g]$.

By the truth-definition in Section 14 we can assume that the domain of g is just the set of variables free in the formulas listed, so that in particular y is not in the domain of g. There are now two cases. The first is that $\mathfrak{A} \vDash \neg(\exists x\phi \wedge \chi_1 \wedge \cdots \wedge \chi_n)[g]$. Then truth-tables show that (130) holds. The second case is that $\mathfrak{A} \vDash (\exists x\phi \wedge \chi_1 \wedge \cdots \wedge \chi_n)[g]$, so there is an element α of $\mathfrak{A}$ such that $\mathfrak{A} \vDash (\phi[y/x] \wedge \chi_1 \wedge \cdots \wedge \chi_n \to \psi)[g, \alpha/y]$, so $\mathfrak{A} \vDash \psi[g, \alpha/y]$. But then since y is not free in ψ, $\mathfrak{A} \vDash \psi[g]$, which again implies (130).

I do not think this solves all the philosophical problems raised by (129). Wiredu [1973] seems relevant.

The references given for the proof calculi discussed in Section 7 remain relevant, except Łukasiewicz and Tarski [1930] which is only about propositional logic. The various theorems of Gentzen [1934], including the cut-elimination theorem, all apply to predicate logic. From the point of view of these calculi, the difference between propositional and predicate logic is relatively slight and has to do with checking that certain symbols don't occur in the wrong places in proofs.

Proof calculi for many-sorted languages are also not hard to come by. See [Schmidt, 1938; Wang, 1952; Feferman, 1968a].

Quantifiers did provoke one quite new proof-theoretic contrivance. In the 1920s a number of logicians (notably Skolem, Hilbert, Herbrand) regarded quantifiers as an intrusion of infinity into the finite-minded world of propositional logic, and they tried various ways of—so to say—deactivating quantifiers. Hilbert proposed the following: replace $\exists x\phi$ everywhere by the sentence $\phi[\varepsilon x\phi/x]$, where '$\varepsilon x\phi$' is interpreted as 'the element I choose among those that satisfy ϕ'. The interpretation is of course outrageous, but Hilbert showed that his ε-calculus proved exactly the same sequents as more conventional calculi. See Hilbert and Bernays [1939] and Leisenring [1969].

It can easily be checked that any sequent which can be proved by the natural deduction calculus sketched above (cf. Sundholm's Chapter in a following volume of this *Handbook* for details) is correct. But nobody could claim to see, just by staring at it, that this calculus can prove *every* correct sequent of predicate logic. Nevertheless it can, as the next section will show.

16 CREATING MODELS

The natural deduction calculus for first-order logic is *complete* in the sense that if $\Delta \models \psi$ then the calculus gives a proof of ψ from assumptions in Δ. This result, or rather the same result for an equivalent Hilbert-style calculus, was first proved by Kurt Gödel in his doctoral dissertation [1930]. Strictly Thoralf Skolem had already proved it in his brilliant papers [1922; 1928; 1929], but he was blissfully unaware that he had done so. (See [Vaught, 1974; Wang, 1970]; Skolem's finitist philosophical leanings seem to have blinded him to some mathematical implications of his work.)

A theory Δ in the language L is said to be *consistent* for a particular proof calculus if the calculus gives no proof of $\bot$ from assumptions in Δ. (Some writers say instead: 'gives no proof of a contradiction $\phi \wedge \neg\phi$ from assumptions in Δ'. For the calculi we are considering, this amounts to the same thing.) We shall demonstrate that *if* Δ *is consistent for the natural deduction calculus then* Δ *has a model.* This implies that the calculus is complete, as follows. Suppose $\Delta \models \psi$. Then $\Delta, \psi \to \bot \models \bot$ (cf. Theorem 4 in Section 5), hence Δ together with $\psi \to \bot$ has no model. But then the theory consisting of Δ together with $\psi \to \bot$ is not consistent for the natural deduction calculus, so we have a proof of $\bot$ from $\psi \to \bot$ and sentences in Δ. One can then quickly construct a proof of ψ from sentences in Δ by the rule (69) for $\bot$.

So the main problem is to show that every consistent theory has a model. This involves constructing a model—but out of what? Spontaneous creation is not allowed in mathematics; the pieces must come from somewhere. Skolem [1922] and Gödel [1930] made their models out of natural numbers, using an informal induction to define the relations. A much more direct source of materials was noticed by Henkin [1949] and independently by Rasiowa and Sikorski [1950]: they constructed the model of Δ out of the theory Δ itself. (Their proof was closely related to Kronecker's [1882] method of constructing extension fields of a field K out of polynomials over K. Both he and they factored out a maximal ideal in a ring.)

Hintikka [1955] and Schütte [1956] extracted the essentials of the Henkin–Rasiowa–Sikorski proof in an elegant form, and what follows is based on their account. For simplicity we assume that the language L has infinitely many individual constants but its only truth-functors are $\neg$ and $\wedge$ and its only quantifier symbol is $\exists$. A theory Δ in L is called a *Hintikka set* if it satisfies these seven conditions:

1. $\bot$ is not in Δ.
2. If ϕ is an atomic formula in Δ then $\neg\phi$ is not in Δ.
3. If $\neg\neg\psi$ is in Δ then ψ is in Δ.
4. If $\psi \wedge \chi$ is in Δ then ψ and χ are both in Δ.

5. If $\neg(\psi \wedge \chi)$ is in Δ then either $\neg\psi$ is in Δ or $\neg\chi$ is in Δ.

6. If $\exists x\psi$ is in Δ then $\psi[c/x]$ is in Δ for some individual constant c.

7. If $\neg\exists x\psi$ is in Δ then $\neg\psi[c/x]$ is in Δ for each individual constant c.

We can construct an L-structure $\mathfrak{A}$ out of a theory Δ as follows. The elements of $\mathfrak{A}$ are the individual constants of L. For each constant $c, I_{\mathfrak{A}}(c)$ is c itself. For each n-place predicate constant R of L the relation $I_{\mathfrak{A}}(R)$ is defined to be the set of all ordered n-tuples $\langle c_1, \ldots, c_n\rangle$ such that the sentence $R(c_1, \ldots, c_n)$ is in Δ.

Let Δ be a Hintikka set. We claim that the structure $\mathfrak{A}$ built out of Δ is a model of Δ. It suffices to show the following, by induction on the complexity of ϕ: if ϕ is in Δ then ϕ is true in $\mathfrak{A}$, and if $\neg\phi$ is in Δ then $\neg\phi$ is true in $\mathfrak{A}$. I consider two sample cases. First let ϕ be atomic. If ϕ is in Δ then the construction of $\mathfrak{A}$ guarantees that $\mathfrak{A} \models \phi$. If $\neg\phi$ is in Δ, then by clause (2), ϕ is not in Δ; so by the construction of $\mathfrak{A}$ again, $\mathfrak{A}$ is not a model of ϕ and hence $\mathfrak{A} \models \neg\phi$. Next suppose ϕ is $\psi \wedge \chi$. If ϕ is in Δ, then by clause (4), both ψ and χ are in Δ; since they have lower complexities than ϕ, we infer that $\mathfrak{A} \models \psi$ and $\mathfrak{A} \models \chi$; so again $\mathfrak{A} \models \phi$. If $\neg\phi$ is in Δ then by clause (5) either $\neg\psi$ is in Δ or $\neg\chi$ is in Δ; suppose the former. Since ψ has lower complexity than ϕ, we have $\mathfrak{A} \models \neg\psi$; it follows again that $\mathfrak{A} \models \neg\phi$. The remaining cases are similar. So *every Hintikka set has a model.*

It remains to show that if Δ is consistent, then by adding sentences to Δ we can get a Hintikka set Δ^+; Δ^+ will then have a model, which must also be a model of Δ because Δ^+ includes Δ. The strategy is as follows.

Step 1. Extend the language L of T to a language L^+ which has infinitely many new individual constants $c_0, c_1, c_2, \ldots$. These new constants are known as the *witnesses* (because in (6) above they will serve as witnesses to the truth of $\exists x\psi$).

Step 2. List all the sentences of L^+ as $\phi_0, \phi_1, \ldots$ in an infinite list so that every sentence occurs infinitely often in the list. This can be done by some kind of zigzagging back and forth.

Step 3. At this very last step there is a parting of the ways. Three different arguments will lead us home. Let me describe them and then compare them.

The first argument we may call the *direct* argument: we simply add sentences to Δ as required by (3)–(7), making sure as we do so that (1) and (2) are not violated. To spell out the details, we define by induction theories $\Delta_0, \Delta_1, \ldots$ in the language L^+ so that (i) every theory Δ_i is consistent; (ii) for all i, Δ_{i+1} includes Δ_i; (iii) for each i, only finitely many of the witnesses appear in the sentences in Δ_i; (iv) Δ_0 is Δ; and (v) for each i, if ϕ_i is in Δ_i then:

3′ if ϕ_i is of form $\neg\neg\psi$ then Δ_{i+1} is Δ_i together with ψ;

4′ if ϕ_i is of form $\psi \wedge \chi$ then Δ_{i+1} is Δ_i together with ψ and χ;

5′ if ϕ_i is of form $\neg(\psi \wedge \chi)$ then Δ_{i+1} is Δ_i together with at least one of $\neg\psi, \neg\chi$;

6′ if ϕ_i is of form $\exists x\psi$ then Δ_{i+1} is Δ_i together with $\psi[c/x]$ for some witness c which doesn't occur in Δ_i;

7′ if ϕ_i is of form $\neg\exists x\psi$ then Δ_{i+1} is Δ_i together with $\neg\psi[c/x]$ for the first witness c such that $\neg\psi[c/x]$ is not already in Δ_i.

It has to be shown that theories Δ_i exist meeting conditions (1)–(5). The proof is by induction. We satisfy (1)–(5) for Δ_0 by putting $\Delta_0 = \Delta$ (and this is the point where we use the assumption that Δ is consistent for natural deduction). Then we must show that if we have got as far as Δ_i safely, Δ_{i+1} can be constructed too. Conditions (2) and (3) are actually implied by the others and (4) is guaranteed from the beginning. So we merely need to show that

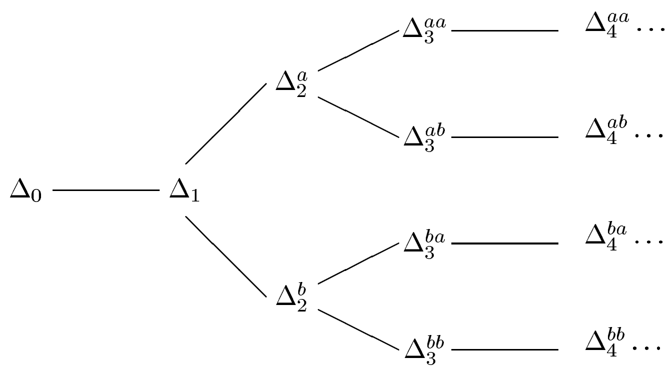
(131) assuming Δ_i is consistent, Δ_{i+1} can be chosen so that it is consistent and satisfies the appropriate one of (3′)– (7′).

There are five cases to consider. Let me take the hardest, which is (6′). It is assumed that ϕ_i is $\exists x\psi$ and is in Δ_i. By (3) so far, some witness has not yet been used; let c be the first such witness and let Δ_{i+1} be Δ_i together with $\psi[c/x]$. If by misfortune Δ_{i+1} was inconsistent, then since c never occurs in Δ_i or ϕ_i, the elimination rule for $\exists$ (section 15 or Sundholm, Volume 2 of this *Handbook*) shows that we can prove $\perp$ already from $\exists x\psi$ and assumptions in Δ_i. But $\exists x\psi$ was in Δ_i, so we have a contradiction to our assumption that Δ_i was consistent. Hence Δ_{i+1} is consistent as required.

When the theories Δ_i have been constructed, let Δ^+ be the set of all sentences which are in at least one theory Δ_i. Since each Δ_i was consistent, Δ^+ satisfies conditions (1) and (2) for a Hintikka set. The requirements (3′)–(7′), and the fact that in the listing $\phi_0, \phi_1, \ldots$ we keep coming round to each sentence infinitely often, ensure that Δ^+ satisfies conditions (3)–(7) as well. So Δ^+ is a Hintikka set and has a model, which completes the construction of a model of Δ.

The second argument we may call the *tree* argument. A hint of it is in [Skolem, 1929]. We imagine a man constructing the theories Δ_i as in the direct argument above. When he faces clauses (3′), (4′), (6′) or (7′), he knows at once how he should construct Δ_{i+1} out of Δ_i; the hardest thing he has to do is to work out which is the first witness not yet used in Δ_i in the case of clause (6′). But in (5′) we can only prove for him that at least one of $\neg\psi$ and $\neg\chi$ can consistently be added to Δ_i, so he must check for himself whether Δ_i together with $\neg\psi$ is in fact consistent. Let us imagine that he is allergic to consistency calculations. Then the best he can do is to make *two alternative suggestions* for Δ_{i+1}, viz. Δ_i with $\neg\psi$, and Δ_i with $\neg\chi$. Thus he will make not a chain of theories $\Delta_0, \Delta_1, \ldots$ but a branching tree of theories:

(132)



Now he no longer knows which of these theories are consistent. So he forgets about consistency and looks directly at conditions (1) and (2) in the definition of a Hintikka set. At least he can tell by inspection whether a theory violates these. So he prunes off the tree all theories which fail (1) or (2)—he can do this as he goes along. Some theories in the tree will become dead ends. But the argument we gave for the earlier direct approach shows that at every level in the tree there must be some theory which can be extended to the next level.

Now a combinatorial theorem known as *König's tree lemma* says that if a tree has a positive but finite number of items at the nth level, for every natural number n, then the tree has a branch which runs up through all these levels. So we know that (132) has an infinite branch. Let $\Delta_0, \Delta_1, \Delta_2, \ldots$ be such a branch and let Δ^+ be the set of all sentences which occur in at least one theory Δ_i in the branch. The previous argument shows that Δ^+ satisfies (3)–(7), and we know that Δ^+ satisfies (1) and (2) because otherwise it would have been pruned off at some finite stage. So again Δ^+ is a Hintikka set.

The third argument is the *maximising* argument, sometimes known as the *Henkin-style* argument, though Skolem's argument in [1922] seems to be of this type. This argument is an opposite to the second kind of argument: instead of using (1)–(7) in the construction and forgetting consistency, we exploit consistency and leave (1)–(7) on one side until the very end. We define by induction theories $\Delta_0, \Delta_1, \ldots$ in the language L^+ so that (i) every theory Δ_i is consistent; (ii) for all i, Δ_{i+1} includes Δ_i; (iii) for each i, only finitely many of the witnesses appear in the sentences in Δ_i; (iv) Δ_0 is Δ; and (v) for each i,

(α) if Δ_i together with ϕ_i is consistent then Δ_{i+1} contains ϕ_i;

(β) if ϕ_i is in Δ_{i+1} and is of form $\exists x\psi$, then for some witness c which doesn't occur in Δ_i or in $\phi_i, \psi[c/x]$ is in Δ_{i+1}.

The argument to justify this construction is the same as for the direct argument, except that (3′), (4′), (5′) and (7′) are now irrelevant. As before, let Δ^+ be the set of sentences which occur in at least one theory Δ_i. Clause (α) in the construction guarantees that

(133) for every sentence ϕ of L^+, if Δ^+ together with ϕ is consistent, then ϕ is in Δ^+.

From (133) and properties of natural deduction we infer

(134) for every sentence ϕ of L^+, if ϕ is provable from assumptions in Δ^+ then ϕ is in Δ^+.

Knowing (133) and (134), we can show that Δ^+ satisfies (3)–(7). For example, take (5) and suppose that $\neg(\psi \wedge \chi)$ is in Δ^+ but $\neg\psi$ is not in Δ^+. Then by (133) there is a proof of $\perp$ from Δ^+ and $\neg\psi$. Using the natural deduction rules we can adapt this proof to get a proof of $\neg\chi$ from Δ^+, and it follows by (134) that $\neg\chi$ is in Δ^+. Since the Δ_i are all consistent, Δ^+ also satisfies (1) and (2). So once again Δ^+ is a Hintikka set.

Some authors take care of clause (β) before the main construction. They can do it by adding to Δ a collection of sentences of the form $\exists x\psi \to \psi[c/x]$. The argument which justified (6′) will justify this too.

The first and third arguments above are very closely related. I gave both of them in the form that would serve for a countable language, but they adapt to first-order languages of any cardinality. The merit of the maximising argument is that the construction is easy to describe. (For example, the listing $\phi_0, \phi_1, \ldots$ need not repeat any formulas.)

The first and second arguments have one advantage over the third. Suppose Δ is a finite set of prenex sentences of form $\exists\vec{x}\forall\vec{y}\psi$, with no quantifiers in ψ. Then these two arguments find Δ^+ after only a finite number of steps in the construction. So Δ^+ is finite and has a finite model, and it follows that we can compute whether or not a sentence of this form has a model. (This is no longer true if function symbols are added to the language as in Section 18 below.) The decidability of propositional logic is a special case of this. So also are various theorems about finite models for modal logics.

When Δ_0 is finite, closer inspection of the trees (132) shows that they are just the natural extension to predicate logic of the semantic tableaux of propositional logic. If Δ_0 has no models then every branch comes to a dead end after a finite number of steps. If Δ_0 has a model, then the tree has a branch which never closes, and we can read this branch as a description of a model. So the tree argument has given us a complete proof calculus for predicate logic. (Cf. Beth [1955; 1962], Jeffrey [1967], Smullyan [1968], Bell and Machover [1977] for predicate logic semantic tableaux.) Incidentally it is most unpleasant to prove the completeness of semantic tableaux by the direct or maximising arguments. One needs facts of the form: if $\Delta \vdash \psi$ and $\Delta, \psi \vdash \chi$ then $\Delta \vdash \chi$. To prove these is to prove Gentzen's cut-elimination theorem.

Notice that even when Δ_0 is finite, semantic tableaux no longer provide a method for deciding whether Δ_0 has a model. If it does have a model, the tree may simply go on branching forever, and we may never know whether it is going to close off in the next minute or the next century. In Section 24 below we prove a theorem of Church [1936] which says that there is not and cannot be any mechanical method for deciding which sentences of predicate logic have models.

17 CONSEQUENCES OF THE CONSTRUCTION OF MODELS

Many of the most important consequences of the construction in the previous section are got by making some changes in the details. For example, instead of using the individual constants of the language as elements, we can number these constants as $b_0, b_1, \ldots$, and use the number n in place of the constant b_n. Since numbers can be thought of as pure sets ([Mendelson, 1987, pp. 187 ff.] or Appendix C below), the structure which emerges at the end will be a pure set structure. Hence, for any standard proof calculus for a language L of predicate logic:

THEOREM 10. *Suppose T is a theory and ψ a sentence of* L*, such that the calculus doesn't prove ψ from T. Then there is a pure set structure which is a model of T and not of ψ.*

In terms of the discussion in Section 8 above, this shows that the Proof Theorist's notion of logical implication agrees with the Model Theorist's, whether or not the Model Theorist restricts himself to pure set structures.

We can take matters one step further by encoding all symbols and formulas of L as natural numbers. So a theory in L will be a set of numbers. Suppose the theory T is in fact the set of all numbers which satisfy the first-order formula ϕ in the language of arithmetic; then by analysing the proof of Theorem 10 we can find another first-order formula χ in the language of arithmetic, which defines a structure with natural numbers as its elements, so that:

THEOREM 11. *In first-order Peano arithmetic we can prove that if some standard proof calculus doesn't prove T is inconsistent, then the structure defined by χ is a model of T.*

(Cf. [Kleene, 1952, p. 394] and [Hasenjaeger, 1953] for a sharper result.)

Theorem 11 is philosophically very interesting. Suppose T is a finite theory, and proof-theoretically T doesn't imply ψ. Applying Theorem 11 to the theory $T \cup \{\neg\psi\}$, we get a formula χ which defines a natural number structure $\mathfrak{A}$ in which T is true and ψ is false. By means of χ, the formulas of T and ψ can be read as meaningful statements about $\mathfrak{A}$ and hence about the natural numbers. The statements in T are true but ψ is false, so we have found an invalid argument of the form 'T. Therefore ψ'. It follows that if a first-order sequent is correct by the Traditional Logician's definition, then it is correct by the Proof Theorist's too. Since the converse is straightforward to prove, we have a demonstration that *the Traditional Logician's notion of validity exactly coincides with the Proof Theorist's.* The proof of this result uses nothing stronger than the assumption that the axioms of first-order Peano arithmetic have a model.

The Traditional Logician's notion of logical implication is quite informal—on any version it involves the imprecise notion of a 'valid English argument'. Nevertheless we have now proved that it agrees exactly with the mathematically precise notion of logical implication given by the Proof Theorist. (Cf. [Kreisel, 1967].) People are apt to say that it is impossible to prove that an informal notion and a formal one agree exactly. Since we have just done the impossible, maybe I should add a comment. Although the notion of a valid argument is vague, there is no doubt that (i) if there is a formal proof of a sequent, then any argument with the form of that sequent must be valid, and (ii) if there is an explicitly definable counterexample to the sequent, then there is an invalid argument of that form. We have shown, by strict mathematics, that every finite sequent has either a formal proof or an explicitly definable counterexample. So we have trapped the informal notion between two formal ones. Contrast *Church's thesis*, that the effectively computable functions (informal notion) are exactly the recursive ones (formal). There is no doubt that the existence of a recursive definition for a function makes the function effectively computable. But nobody has yet thought of any kind of mathematical object whose existence undeniably implies that a function is *not* effectively computable. So Church's thesis remains unproved. (Van Dalen's chapter in this Volume discusses Church's thesis.)

I return to the completeness proof. By coding all expressions of L into numbers or sets, we made it completely irrelevant that the symbols of L can be written on a page, or even that there are at most countably many of them. *So let us now allow arbitrary sets to serve instead of symbols.* Languages of this abstract type can be called *set languages*. They are in common use today even among proof theorists. Of course to use these languages we have to rely either on our intuitions about sets or on proofs in axiomatic set theory; there is no question of checking by inspection. Henkin's [1949] completeness proof was given in this setting. In fact he proved:

THEOREM 12. *If* L *is a first-order set language and T a theory in* L *whose cardinality is at most the infinite cardinal κ, then either a finite part of T can be proved inconsistent by a proof calculus, or T has a model with at most κ elements.*

Theorem 12 has several important mathematical consequences. For example, the *Compactness Theorem* says:

THEOREM 13. *Let T be a first-order theory (in a set language). If every finite set of sentences in T has a model, then T has a model.*

Theorem 13 for countable languages was proved by Gödel in [1930]. For propositional logic with arbitrarily many symbols it was proved by Gödel [1931a], in answer to a question of Menger. The first proof of Theorem 13 was sketched rather inadequately by Anatolii Mal'tsev in [1936] (see the review of [Mal'tsev, 1941] by Henkin and Mostowski [1959]). But in [1941] Mal'tsev showed that Theorem 13 has interesting and far from trivial consequences in group theory, thus beginning one of the most important lines of application of first-order logic in mathematics.

The last consequence I shall draw from Theorem 12 is not really interesting until identity is added to the language (see the next section); but this is a convenient place to state it. It is the *Upward and Downward Löwenheim–Skolem Theorem*:

THEOREM 14. *Let T be a first-order theory in a language with λ formulas, and κ an infinite cardinal at least as great as λ. If T has a model with infinitely many elements then T has one with exactly κ elements.*

Theorem 13 was proved in successively stronger versions by Löwenheim [1915], Skolem [1920; 1922], Tarski in unpublished lectures in 1928, Mal'tsev [1936] and Tarski and Vaught [1956]; see [Vaught, 1974] for a thorough history of this and Theorems 12 and 13. The texts of Bell and Slomson [1969], Chang and Keisler [1973] and Hodges [1993a] develop these theorems, and Sacks [1972] and Cherlin [1976] study some of their applications in algebra. Skolem [1955] expressly dissociated himself from the Upward version of Theorem 14, which he regarded as nonsense.

18 IDENTITY

The symbol '=' is reserved for use as a 2-place predicate symbol with the intended meaning

(135) $a = b$ iff a and b are one and the same thing.

When $\mathfrak{A}$ is a structure for a language containing '=', we say that $\mathfrak{A}$ has *standard identity* if the relation $I_{\mathfrak{A}}(=)$ holds between elements α and β of $\mathfrak{A}$ precisely when α and β are the same element.

'$x = y$' is read as 'x equals y', rather misleadingly—all men may be created equal but they are not created one and the same man. Another reading is 'x is identical with y'. As far as English usage goes, this is not much improvement on 'equals': there are two identical birds feeding outside my window, but they aren't the same bird (and think of identical twins). Be that as it may, '=' is called the *identity* sign and the relation it expresses in (135) is called *identity*.

Let L be a language containing the symbol '='. It would be pleasant if we could find a theory Δ in L whose models are exactly the L-structures with standard identity. Alas, there is no such theory. *For every* L-*structure* $\mathfrak{A}$ *with standard identity there is an* L-*structure* $\mathfrak{B}$ *which is a model of the same sentences of* L *as* $\mathfrak{A}$ *but doesn't have standard identity.* Let us prove this.

Take an L-structure $\mathfrak{A}$ with standard identity and let $\delta_1, \ldots, \delta_{2,000,000}$ be two million objects which are not in the domain of $\mathfrak{A}$. Let β be an element of $\mathfrak{A}$. We construct the L-structure $\mathfrak{B}$ thus. The elements of $\mathfrak{B}$ are those of $\mathfrak{A}$ together with $\delta_1, \ldots, \delta_{2,000,000}$. For each individual constant c we put $I_{\mathfrak{B}}(c) = I_{\mathfrak{A}}(c)$. For each element α of $\mathfrak{B}$ we define an element $\hat{\alpha}$ of $\mathfrak{A}$ as follows: if α is in the domain of $\mathfrak{A}$ then $\hat{\alpha}$ is α, and if α is one of the δ_j's then $\hat{\alpha}$ is β. For every n-place predicate constant R we choose $I_{\mathfrak{B}}(R)$ so that if $\langle \alpha_1, \ldots, \alpha_n \rangle$ is any n-tuple of elements of $\mathfrak{B}$, then:

(136) $\langle \alpha_1, \ldots, \alpha_n \rangle$ is in $I_{\mathfrak{B}}(R)$ iff $\langle \hat{\alpha}_1, \ldots, \hat{\alpha}_n \rangle$ is in $I_{\mathfrak{A}}(R)$.

This defines $\mathfrak{B}$. By induction on the complexity of ϕ we can prove that for every formula $\phi(x_1, \ldots, x_n)$ of L and every n-tuple $\langle \alpha_1, \ldots, \alpha_n \rangle$ of elements of $\mathfrak{B}$,

(137) $\mathfrak{B} \vDash \phi[\alpha_1/x_1, \ldots, \alpha_n/x_n]$ iff $\mathfrak{A} \vDash \phi[\hat{\alpha}_1/x_1, \ldots, \hat{\alpha}_n/x_n]$.

In particular $\mathfrak{A}$ and $\mathfrak{B}$ are models of exactly the same sentences of L. Since $\mathfrak{A}$ has standard identity, $\mathfrak{A} \vDash (x = x)[\beta/x]$. Then from (136) it follows that the relation $I_{\mathfrak{B}}(=)$ holds between any two of the elements $\delta_1, \ldots, \delta_{2,000,000}$, and so $I_{\mathfrak{B}}(=)$ is vastly different from standard identity.

So we look for a second best. Is there a theory Δ which is true in all L-structures with standard identity, and which logically implies every sentence of L that is true in all such L-structures? This time the answer is positive. The following theory will do the job:

(138) $\forall x\ x = x$.

(139) All sentences of the form $\forall\vec{z}xy(x = y \to (\phi \to \phi[y/x]))$.

Formula (138) is known as the *law of reflexivity of identity.* (139) is not a single sentence but an infinite family of sentences, namely all those which can be got by putting any formula ϕ of L into the expression in (139); $\vec{z}$ are all the free variables of ϕ except for x and y. These sentences (139) are collectively known as *Leibniz' Law.* They are the nearest we can get within L to saying that if $a = b$ then anything which is true of a is true of b too.

By inspection it is clear that every L-structure with standard identity is a model of (138) and (139). To show that (138) and (139) logically imply every sentence true in all structures with standard identity, let me prove something stronger, namely: *For every* L-*structure* $\mathfrak{B}$ *which is a model of* (138) *and* (139) *there is an* L-*structure* $\mathfrak{A}$ *which is a model of exactly the same sentences of* L *as* $\mathfrak{B}$ *and has standard identity.* Supposing this has been proved, let Δ be the theory consisting of (138) and (139), and let ψ be a sentence of L which is not logically implied by Δ. Then some L-structure $\mathfrak{B}$ is a model of Δ and $\neg\psi$; so some structure $\mathfrak{A}$ with standard identity is also a model of $\neg\psi$. It follows that ψ is not true in all structures with standard identity.

To prove what I undertook to prove, let $\mathfrak{B}$ be a model of Δ. Then we can show that the following hold, where we write $=_{\mathfrak{B}}$ for $I_{\mathfrak{B}}(=)$:

(140) the relation $I_{\mathfrak{B}}(=)$ is an equivalence relation;

(141) for every n-place predicate constant R of L, if $\alpha_1 =_{\mathfrak{B}} \beta_1, \dots, \alpha_n =_{\mathfrak{B}} \beta_n$ and $\langle\alpha_1, \dots, \alpha_n\rangle$ is in $I_{\mathfrak{B}}(R)$ then $\langle\beta_1, \dots, \beta_n\rangle$ is in $I_{\mathfrak{B}}(R)$.

Statement (141) can be proved by applying Leibniz' Law n times. Then (140) follows from (141) and reflexivity of identity, taking '=' for R. Statements (140) and (141) together are summarised by saying that the relation $=_{\mathfrak{B}}$ is a *congruence* for L. For each element α of $\mathfrak{B}$, we write $\alpha^=$ for the equivalence class of α under the relation $=_{\mathfrak{B}}$.

Now we define the L-structure $\mathfrak{A}$ as follows. The domain of $\mathfrak{A}$ is the class of all equivalence classes $\alpha^=$ of elements α of $\mathfrak{B}$. For each individual constant c we define $I_{\mathfrak{A}}(c)$ to be $I_{\mathfrak{B}}(c)^=$. For each n-place predicate symbol R of L we define $I_{\mathfrak{A}}(R)$ by:

(142) $\langle\alpha_1^=, \dots, \alpha_n^=\rangle$ is in $I_{\mathfrak{A}}(R)$ iff $\langle\alpha_1, \dots, \alpha_n\rangle$ is in $I_{\mathfrak{B}}(R)$.

Definition (142) presupposes that the right-hand side of (142) is true or false depending only on the equivalence classes of $\alpha_1, \dots, \alpha_n$; but (141) assured this.

In particular, $\alpha^= =_{\mathfrak{A}} \beta^=$ if and only if $\alpha =_{\mathfrak{B}} \beta$, in other words, if and only if $\alpha^=$ equals $\beta^=$. Hence, $\mathfrak{A}$ has standard identity. It remains only to show that for every formula $\phi(x_1, \dots, x_n)$ of L and all elements $\alpha_1, \dots, \alpha_n$ of $\mathfrak{B}$,

(143) $\mathfrak{A} \vDash \phi[\alpha_1^{=}/x_1, \ldots, \alpha_n^{=}/x_n]$ iff $\mathfrak{B} \vDash \phi[\alpha_1/x_1, \ldots, \alpha_n/x_n]$.

Statement (143) is proved by induction on the complexity of ϕ.

Most logicians include '=' as part of the vocabulary of every language for predicate logic, and interpret it always to mean standard identity. Since it is in every language, it is usually not mentioned in the similarity type. The proof calculi have to be extended to accommodate '='. One way to extend the natural deduction calculus is to add two new rules:

(144) $$\frac{}{x = x} \qquad \frac{x = y \qquad \phi}{\phi[y/x]}$$

The first rule deduces $x = x$ from no premises.

Identity is needed for virtually all mathematical applications of logic. It also makes it possible to express in formulas the meanings of various English phrases such as 'the', 'only', 'at least one', 'at most eight', etc. (see e.g. Section 21 below).

Many mathematical applications of logic need symbols of another kind, called *function symbols*. The definitions given above can be stretched to allow function symbols as follows. Symbols f, g, h etc., with or without subscripts, are called *function constants*. A similarity type may contain function constants, each of which is labelled as an *n-place constant* for some positive integer n. If the language L has an n-place function constant f and $\mathfrak{A}$ is an L-structure, then f is interpreted by $\mathfrak{A}$ as an *n-place function* $I_{\mathfrak{A}}(f)$ which assigns one element of $\mathfrak{A}$ to each ordered n-tuple of elements of $\mathfrak{A}$. For example the 2-place function constant '+' may be interpreted as a function which assigns 5 to $\langle 2, 3\rangle$, 18 to $\langle 9, 9\rangle$ and so forth—though of course it can also be interpreted as some quite different function.

There are various ways of writing functions, such as

(145) $\sin x, \sqrt{x}, x^2, \hat{x}, y^y, x + y, \langle x, y\rangle$.

But the general style is '$f(x_1, \ldots, x_n)$', and logicians' notation tends to follow this style. The details of syntax and proof theory with function symbols are rather messy, so I omit them and refer the reader to [Hilbert and Bernays, 1934] for details.

One rarely needs function symbols outside mathematical contexts. In any case, provided we have '=' in our language, everything that can be said with function symbols can also be said without them. Briefly, the idea is to use a predicate constant R in such a way that '$R(x_1, \ldots, x_{n+1})$' means '$f(x_1, \ldots, x_n) = x_{n+1}$'. When the function symbol f is in the language, it is true in all structures—and hence logically valid—that for all $x_1, \ldots, x_n$ there is a unique x_{n+1} such that $f(x_1, \ldots, x_n) = x_{n+1}$. Translating f into R, this becomes

(146) $\forall x_1 \cdots x_n z t \exists y((R(x_1, \ldots, x_n, z) \wedge R(x_1, \ldots, x_n, t) \rightarrow z = t) \wedge R(x_1, \ldots, x_n, y))$.

Since (146) is not logically valid, it may have to be assumed as an extra premise when we translate arguments involving f into arguments involving R.

19 AXIOMS AS DEFINITIONS

Axioms are, roughly speaking, the statements which one writes down at the beginning of a book in order to define the subject-matter of the book and provide a basis for deductions made in the book. For example any textbook of group theory will start by telling you that a group is a triple $\langle G, *, e\rangle$ where $*$ is a binary operation in the set G and e is an element of G such that

(147) $*$ is associative, i.e. for all x, y and $z, x * (y * z) = (x * y) * z$,

(148) e is an identity, i.e. for all $x, x * e = e * x = x$,

(149) every element x has an inverse, i.e. an element y such that $x*y = y * x = e$.

Statements (147)–(149) are known as the *axioms for groups*. I could have chosen examples from physics, economics or even ethics.

It is often said that in an 'axiomatic theory' such as group theory, the axioms are 'assumed' and the remaining results are 'deduced from the axioms'. This is completely wrong. W. R. Scott's textbook *Group Theory* [1964] contains 457 pages of facts about groups, and the last fact which can by any stretch of the imagination be described as being 'deduced from (147)–(149)' occurs on page 8. We could indeed rewrite Scott's book as a set of deductions from assumed axioms, but the axioms would be those of set theory, not (147)–(149). These three group axioms would appear, not as assumptions but as *part of the definition of 'group'*.

The definition of a group can be paraphrased as follows. First we can recast the triple $\langle G, *, e\rangle$ as an L-structure $\mathfrak{G} = \langle G, I_{\mathfrak{G}}\rangle$ in a first-order language L with one 2-place function symbol $*$ and one individual constant e. Then $\mathfrak{G}$ is a group if and only if $\mathfrak{G}$ is a model of the following three sentences:

(150) $\forall xyz\ x * (y * z) = (x * y) * z$,

(151) $\forall x(x * e = x \wedge e * x = x)$,

(152) $\forall x \exists y(x * y = e \wedge y * x = e)$.

Generalising this, let Δ be any theory in a first-order language L. Let $\mathbf{K}$ be a class of L-structures. Then Δ is said to *axiomatise* $\mathbf{K}$, and $\mathbf{K}$ is called $Mod(\Delta)$, if $\mathbf{K}$ is the class of all L-structures which are models of Δ. The sentences in Δ are called *axioms* for $\mathbf{K}$. Classes of form $Mod(\{\phi\})$ for a single first-order sentence ϕ are said to be *first-order definable*. Classes of form $Mod(\Delta)$ for a first-order theory Δ are said to be *generalised first-order definable*. The class of groups is first-order definable—we can use the conjunction of the three sentences (150)–(152).

Many other classes of structure which appear in pure or applied mathematics are (generalised) first-order definable. To give examples I need only list the axioms. First, *equivalence relations*:

(153) $\forall x R(x,x)$ ‘R is reflexive’

(154) $\forall xy(R(x,y) \to R(y,x))$ ‘R is symmetric’

(155) $\forall xyz(R(x,y) \wedge R(y,z) \to R(x,z))$ ‘R is transitive’.

Next, *partial orderings*:

(156) $\forall x\ x \leq x$ ‘$\leq$ is reflexive’

(157) $\forall xyz(x \leq y \wedge y \leq z \to x \leq z)$ ‘$\leq$ is transitive’

(158) $\forall xy(x \leq y \wedge y \leq x \to x = y)$ ‘$\leq$ is antisymmetric’.

Then *total* or *linear orderings* are axiomatised by (157) and (158) and

(159) $\forall xy(x \leq y \vee y \leq x)$ ‘$\leq$ is connected’.

Total orderings can also be axiomatised as follows, using $<$ instead of $\leq$:

(160) $\forall xyz(x < y \wedge y < z \to x < z)$

(161) $\forall x \neg x < x$

(162) $\forall xy(x < y \vee y < x \vee x = y)$.

A total ordering in the second style can be converted into a total ordering in the first style by reading $x \leq y$ as meaning $x < y \vee x = y$. There is a similar conversion from the first style to the second. We can express various conditions on linear orderings by adding further axioms to (157)–(159):

(163) $\exists x \forall y\ y \leq x$ ‘there is a last element’

(164) $\forall x \exists y(\neg x = y \wedge \forall z(x \leq z \leftrightarrow x = z \vee y \leq z))$
‘every element has an immediate successor’.

Algebra is particularly rich in first-order or generalised first-order definable classes, for example rings, fields, lattices, categories, toposes, algebraically closed fields, vector spaces over a given field. *Commutative groups* are axiomatised by adding to (150)–(152) the axiom

(165) $\forall xy\ x * y = y * x$.

All the examples mentioned so far are first-order definable except for algebraically closed fields and vector spaces over an infinite field, which need infinitely many sentences to define them.

The notion of first-order definable classes was first made explicit in a paper of Tarski [1954]. If we know that a class of structures is generalised first-order definable then we immediately know various other facts about it, for example that it is closed under taking ultraproducts (cf. [Chang and Keisler, 1973] or [Bell and Slomson, 1969]—they are defined in Appendix C below) and that implicit definitions in the class can all be made explicit ('Beth's theorem'—Theorem 33 in Section 27 below). On the other hand, if one is not interested in model-theoretic facts like these, the informal style of (147)–(149) makes just as good a definition of a class as any set of first-order formulas. (In the philosophy of science, structuralists have given reasons for preferring the informal set-theoretic style; see [Sneed, 1971] and [Stegmüller, 1976].)

It was Hilbert and his school who first exploited axioms, higher-order as well as first-order, as a means of defining classes of structures. Hilbert was horrifically inaccurate in describing what he was doing. When he set up geometric axioms, he said that they defined what was meant by a point. Frege then caustically asked how he could use this definition to determine whether his pocket watch was a point ([Frege and Hilbert, 1899–1900]). Hilbert had simply confused defining a class of structures with defining the component relations and elements of a single structure. (Cf. the comments of [Bernays, 1942].) In this matter Hilbert was a spokesman for a confusion which many people shared. Even today one meets hopeful souls who believe that the axioms of set theory define what is meant by 'set'.

Hilbert added the lunatic remark that 'If ... arbitrarily posited axioms together with all their consequences do not contradict one another, then they are true and the things defined by these axioms exist' [Frege and Hilbert, 1899–1900]. For example, one infers, if the axioms which say there is a measurable cardinal are consistent, then there is a measurable cardinal. If the axioms which say there is no measurable cardinal are consistent, then there is no measurable cardinal. If both sets of axioms are consistent .... In later years he was more cautious. In fairness to Hilbert, one should set his remark against the background beliefs of his time, one of which was the now happily discredited theory of 'implicit definition' (nothing to do with Beth's theorem of that name). See [Coffa, 1991], who puts the Frege-Hilbert debate helpfully into a broad historical context. Be that as it may, readers of Hilbert's philosophical remarks should always bear in mind his slogan '*Wir sind Mathematiker*' [Hilbert, 1926].

20 AXIOMS WITH INTENDED MODELS

Axioms are not always intended to define a class of structures as in Section 19 above. Often they are written down *in order to set on record certain facts about a particular structure.* The structure in question is then called the *intended interpretation* or *standard model* of the axioms. The best known example is probably the axioms of Peano arithmetic, which were set down by Dedekind [1888; 1967] as a statement of the 'fundamental properties' of the natural number sequence (the first-order formalisation is due to Gödel [1931b], cf. Appendix B below). Euclid's axioms and postulates of geometry are another example, since he undoubtedly had space in mind as the intended interpretation.

The object in both Dedekind's case and Euclid's was to write down some elementary facts about the standard model so that further information could be got by making deductions from these facts. With this aim it becomes very natural to write the axioms in a first-order language, because we understand first-order deducibility well and so we shall know exactly what we are entitled to deduce from the axioms.

However, there is no hope at all of *defining* the natural numbers, even up to isomorphism, by means of any first-order axioms. Let me sketch a proof of this—it will be useful later. Suppose L is the first-order language of arithmetic, with symbols to represent plus and times, a 2-place predicate constant $<$ ('less than'), and a name n^* for each natural number n. Let L^+ be L with a new individual constant c added. Let Δ be the set of all sentences of L which are true in the standard model. Let Δ^+ be Δ together with the sentences

(166) $0^* < c, \quad 1^* < c, \quad 2^* < c, \ldots.$

Now if Γ is any finite set of sentences from Δ^+ then Γ has a model: take the standard model of Δ and let c stand for some natural number which is greater than every number mentioned in Γ. So by the Compactness Theorem (Theorem 13 in Section 17 above), Δ^+ has a model $\mathfrak{A}$. Since Δ^+ includes Δ, $\mathfrak{A}$ is a model of Δ and hence is a model of exactly the same sentences of L as the standard model. But $\mathfrak{A}$ also has an element $I_{\mathfrak{A}}(c)$ which by (166) is 'greater than' $I_{\mathfrak{A}}(0^*), I_{\mathfrak{A}}(1^*), I_{\mathfrak{A}}(2^*)$ and all the 'natural numbers' of $\mathfrak{A}$. So $\mathfrak{A}$ is a model of Δ with an 'infinite element'. Such models of Δ are called *non-standard models of arithmetic.* They were first constructed by Skolem [1934], and today people hold conferences on them.

But one can reasonably ask whether, say, the first-order Peano axioms (cf. Appendix B) imply all first-order sentences which are true in the standard model. This is equivalent to asking whether the axioms are a *complete* theory in the sense that if ϕ is any sentence of their language, then either ϕ or $\neg\phi$ is a consequence of the axioms. Gödel's epoch-making paper [1931b] showed that the first-order Peano axioms are not complete; in fact no mechanically describable theory in this language is both complete and true in the standard model. In Section 24 below I shall sketch a proof of this.

There is a halfway house between the use of axioms to define a class and their use to say things about a standard model. Often we want to work with a class **K** of L-structures which may not be generalised first-order definable. In such cases we say that a theory Δ is a *set of axioms* for **K** if every structure in **K** is a model of Δ; we call it a *complete* set of axioms for **K** if moreover every sentence of L which is true in all structures in **K** is a logical consequence of Δ.

Let me give three examples. (i) For the first, paraphrasing Carnap [1956, p. 222 ff] I consider the class of all structures which represent possible worlds, with domain the set of all people, 'Bx' interpreted as 'x is a bachelor' and 'Mx' as 'x is married'. Obviously this class is not generalised first-order definable. But the following sentence is a complete set of axioms:

(167) $\forall x(Bx \to \neg Mx)$.

In Carnap's terminology, when **K** is the class of all structures in which certain symbols have certain fixed meanings, axioms for **K** are called *meaning postulates.* (Lakoff [1972] discusses some trade-offs between meaning postulates and deep structure analysis in linguistics.)

(ii) For a second sample, consider second-order logic (cf. [Chapter 4, below]). In this logic we are able to say 'for all subsets P of the domain, ...', using second-order quantifiers '$\forall P$'. For reasons explained in Chapter 4 below, there is no hope of constructing a complete proof calculus for second-order logic. But we do have some incomplete calculi which are good for most practical purposes. They prove, among other things, the formula

(168) $\forall PQ(\forall z(P(z) \leftrightarrow Q(z)) \to P = Q)$

which is the second-order logician's version of the axiom of extensionality.

Second-order logic can be translated wholesale into a kind of two-sorted first-order logic by the following device. Let L be any (first-order) language. Form a two-sorted language $L^{\downarrow}$ with the same predicate and individual constants as L, together with one new 2-place predicate constant ε. For each L-structure $\mathfrak{A}$, form the $L^{\downarrow}$-structure $\mathfrak{A}^{\downarrow}$ as follows. The domain of $\mathfrak{A}^{\downarrow}$ is $|\mathfrak{A}| \cup \mathcal{P}|\mathfrak{A}|$, $|\mathfrak{A}|$ is the domain for the first sort and $\mathcal{P}|\mathfrak{A}|$ is the domain for the second. ($\mathcal{P}X$ = the set of all subsets of X.) If α and β are elements of $\mathfrak{A}^{\downarrow}$, then

(169) $\langle\alpha, \beta\rangle$ is in $I_{\mathfrak{A}^{\downarrow}}(\varepsilon)$ iff α is an element of the first sort, β of the second sort, and $\alpha \in \beta$.

The constants of L are interpreted in the first sort of $\mathfrak{A}^{\downarrow}$ just as they were in $\mathfrak{A}$. Now each second-order statement ϕ about L-structures $\mathfrak{A}$ is equivalent to a first-order statement $\phi^{\downarrow}$ about $L^{\downarrow}$- structures $\mathfrak{A}^{\downarrow}$. For example, if we use number superscripts to distinguish the first and second sorts of variables, the axiom of extensionality (168) translates into

$$(170)\ \ \forall x^2 y^2(\forall z^1(z^1 \varepsilon x^2 \leftrightarrow z^1 \varepsilon y^2) \to x^2 = y^2).$$

Axiom (170) is a first-order sentence in $L^{\downarrow}$.

Let $\mathbf{K}$ be the class of all $L^{\downarrow}$-structures of form $\mathfrak{A}^{\downarrow}$ for some L-structure $\mathfrak{A}$. Let $\mathbf{QC}^2$ be some standard proof calculus for second-order logic, and let Δ be the set of all sentences $\phi^{\downarrow}$ such that ϕ is provable by $\mathbf{QC}^2$. Then Δ is a set of axioms of $\mathbf{K}$, though not a complete one. The $L^{\downarrow}$-structures in $\mathbf{K}$ are known as the *standard* models of Δ. There will be plenty of non-standard models of Δ too, but because of (170) they can all be seen as 'parts of' standard models in the following way. For each element β of the second sort in the model $\mathfrak{B}$ of Δ, let β^+ be the set of elements α such that $\langle\alpha, \beta\rangle \in I_{\mathfrak{B}}(\varepsilon)$. By (170), $\beta^+ = \gamma^+$ implies $\beta = \gamma$. So in $\mathfrak{B}$ we can replace each element β of the second sort by β^+. Then the second sort consists of subsets of the domain of the first sort, but not necessarily all the subsets. All the subsets are in the second domain if and only if this doctored version of $\mathfrak{B}$ is a standard model. (Models of Δ, standard or non-standard, are known as *Henkin models of second-order logic*, in view of [Henkin, 1950].)

How can one distinguish between a proof calculus for second-order logic on the one hand, and on the other hand a first-order proof calculus which also proves the sentences in Δ? The answer is easy: one can't. In our notation above, the proof calculus for second-order logic has '$P(z)$' where the first-order calculus has '$z^1 \varepsilon x^2$', but this is no more than a difference of notation. Take away this difference and the two calculi become exactly the same thing. Don't be misled by texts like Church [1956] which present 'calculi of first order' in one chapter and 'calculi of second order' in another. The latter calculi are certainly different from the former, because they incorporate a certain amount of set theory. But what makes them second-order calculi, as opposed to two-sorted first-order calculi with extra non-logical axioms, is *solely their intended interpretation.*

It follows, incidentally, that it is quite meaningless to ask whether the proof theory of actual mathematics is first-order or higher-order. (I recently saw this question asked. The questioner concluded that the problem is 'not easy'.)

Where then can one meaningfully distinguish second-order from first-order? One place is the *classification of structures.* The class $\mathbf{K}$ of standard models of Δ is not a first-order definable class of $L^{\downarrow}$-structures, but it is second-order definable.

More controversially, we can distinguish between *first-order and second-order statements about a specific structure*, even when there is no question of classification. For example the sentence (168) says about an L-structure

$\mathfrak{A}$ something which can't be expressed in the first-order language of $\mathfrak{A}$. This is not a matter of classification, because (168) is true in *all* L-structures.

(iii) In Section 18 we studied the class of all L-structures with standard identity. Quine [1970, p. 63f] studies them too, and I admire his nerve. He first demonstrates that in any language L with finite similarity type there is a formula ϕ which defines a congruence relation in every L-structure. From Section 18 we know that ϕ cannot always express identity. Never mind, says Quine, let us *redefine* identity by the formula ϕ. This happy redefinition instantly makes identity first-order definable, at least when the similarity type is finite. It also has the consequence, not mentioned by Quine, that for any two different things there is some language in which they are the same thing. (Excuse me for a moment while I redefine exams as things that I don't have to set.)

21 NOUN PHRASES

In this section I want to consider whether we can make any headway by adding to first-order logic some symbols for various types of noun phrase. Some types of noun phrase, such as 'most Xs', are not really fit for formalising because their meanings are too vague or too shifting. Of those which can be formalised, some never give us anything new, in the sense that any formula using a symbol for them is logically equivalent to a formula of first-order logic (with $=$); to express this we say that these formalisations give *conservative extensions* of first-order logic. Conservative extensions are not necessarily a waste of time. Sometimes they enable us to say quickly something that can only be said lengthily in first-order symbols, sometimes they behave more like natural languages than first-order logic does. So they may be useful to linguists or to logicians in a hurry.

Many (perhaps most) English noun phrases have to be symbolised as *quantifiers and not as terms.* For example the English sentence

(171) I have inspected every batch.

finds itself symbolised by something of form

(172) For every batch x, I have inspected x.

Let me recall the reason for this. If we copied English and simply put the noun phrase in place of the variable x, there would be no way of distinguishing between (i) the negation of 'I have inspected every batch' and (ii) the sentence which asserts, of every batch, that I have not inspected it. In style (172) there is no confusion between (i), viz.

(173) $\neg$ For every batch x, I have inspected x.

and (ii), viz.

(174) For every batch x, $\neg$ I have inspected x.

Confusions like that between (i) and (ii) are so disastrous in logic that it constantly amazes logicians to see that natural languages, using style (171), have not yet collapsed into total anarchy.

In the logician's terminology, the *scope* of the quantifier 'For every batch x' in (174) is the whole sentence, while in (173) it is only the part after the negation sign. Unlike its English counterpart, the quantifier doesn't *replace* the free occurrences of x in the predicate, it *binds* them. (More precisely, an occurrence of a quantifier with variable x binds all occurrences of x which are within its scope and not already bound.) This terminology carries over at once to the other kinds of quantifier that we shall consider, for example

(175) $\neg$ For one in every three men x, x is colour blind.

The quantifier 'For one in every three men x' binds both occurrences of the variable, and doesn't include the negation in its scope.

I shall consider three groups of noun phrases. The first yield conservative extensions of first-order logic and are quite unproblematic. The second again give conservative extensions and are awkward. The third don't yield conservative extensions—we shall prove this. In all cases I assume that we start with a first-order language L with identity.

The *first* group are noun phrases such as 'At least n things x such that ϕ'. We do it recursively:

(176) $\exists_{\geq 0} x\phi$ is $\neg\bot$; $\quad \exists_{\geq 1} x\phi$ is $\exists x\phi$.

(177) $\exists_{\geq n+1} x\phi$ is $\exists y(\phi[y/x] \wedge \exists_{\geq n} x(\neg x = y \wedge \phi))$ when $n \geq 1$.

To these definitions we add:

(178) $\exists_{\leq n} x\phi$ is $\neg\exists_{\geq n+1} x\phi$.

(179) $\exists_{=n} x\phi$ is $\exists_{\geq n} x\phi \wedge \exists_{\leq n} x\phi$.

$\exists_{=1} x\phi$ is sometimes written $\exists! x\phi$.

Definitions (176)–(179) are in the metalanguage; they simply select formulas of L. But there is no difficulty at all in adding the symbols $\exists_{\geq n}$, $\exists_{\leq n}$ and $\exists_{=n}$ for each natural number to the language L, and supplying the needed extra clauses in the definition of $\vDash$, together with a complete formal calculus.

The *second* group are singular noun phrases of the form 'The such-and-such'. These are known as *definite descriptions.* Verbal variants of definite descriptions, such as 'My father's beard' for 'The beard of my father', are generally allowed to be definite descriptions too.

According to Bertrand Russell [1905], Whitehead and Russell [1910, Introduction, Chapter III], the sentence

(180) The author of 'Slawkenburgius on Noses' was a poet.

can be paraphrased as stating three things: (1) at least one person wrote 'Slawkenburgius on Noses'; (2) at most one person wrote 'Slawkenburgius on Noses'; (3) some person who did write 'Slawkenburgius on Noses' was a poet. I happily leave to Bencivenga [4.5] and Salmon [8.5] the question whether Russell was right about this. But assuming he was, his theory calls for the following symbolisation. We write '$\{\imath x\psi\}$' to represent 'the person or thing x such that ψ', and we define

(181) $\{\imath x\psi\}\phi$ to mean $\exists_{=1}x\psi \wedge \exists x(\psi \wedge \phi)$.

Expression (181) can be read either as a metalinguistic definition of a formula L, or as a shorthand explanation of how the expressions $\{\imath x\psi\}$ can be added to L. In the latter case the definition of $\models$ has to sprout one extra clause:

(182) $\mathfrak{A} \models \{\imath x\psi\}\phi[g]$ iff there is a unique element α of $\mathfrak{A}$ such that $\mathfrak{A} \models \psi[g, \alpha/x]$, and for this $\alpha, \mathfrak{A} \models \phi[g, \alpha/x]$.

There is something quite strongly counterintuitive about the formulas on either side in (181). It seems in a way obvious that when there is a unique such-and-such, we can refer to it by saying 'the such-and-such'. But Russell's paraphrase never allows us to use the expression $\{\imath x\psi\}$ this way. For example if we want to say that the such-and-such equals 5, Russell will not allow us to render this as '$\{\imath x\psi\} = 5$'. The expression $\{\imath x\psi\}$ has the wrong grammatical type, and the semantical explanation in (182) doesn't make it work like a name. On the right-hand side in (181) the position is even worse—the definition description has vanished without trace.

Leaving intuition on one side, there are any number of places in the course of formal calculation where one wants to be able to say 'the such-and-such', and then operate with this expression *as a term.* For example formal number theorists would be in dire straits if they were forbidden use of the term

(183) $\mu x\psi$, i.e. the least number x such that ψ.

Likewise formal set theorists need a term

(184) $\{x|\psi\}$, i.e. the set of all sets x such that ψ.

Less urgently, there are a number of mathematical terms which bind variables, for example the integral $\int_b^a f(x)dx$ with bound variable x, which are naturally defined as 'the number λ such that ...(here follows half a page of calculus)'. If we are concerned to formalise mathematics, the straightforward way to formalise such an integral is by a definite description term.

Necessity breeds invention, and in the event it is quite easy to extend the first-order language L by adding *terms* $\imath x\psi$. (The definitions of 'term' and 'formula' in Section 13 above have to be rewritten so that the classes are defined by simultaneous induction, because now we can form terms out of formulas as well as forming formulas out of terms.) There are two ways to proceed. One is to take $\imath x\psi$ as a name of the unique element satisfying ψ, if there is such a unique element, and as undefined otherwise; then to reckon an atomic formula false whenever it contains an undefined term. This is equivalent to giving each occurrence of $\imath x\psi$ the smallest possible scope, so that the notation need not indicate any scope. (Cf. [Kleene, 1952, p. 327]; [Kalish and Montague, 1964, Chapter VII].) The second is to note that questions of scope only arise if there is not a unique such-and-such. So we can choose a constant of the language, say 0, and read $\imath x\psi$ as

(185) the element which is equal to the unique x such that ψ if there is such a unique x, and is equal to 0 if there is not.

(Cf. [Montague and Vaught, 1959; Suppes, 1972].)

Russell himself claimed to believe that definite descriptions 'do not name'. So it is curious to note (as Kaplan does in his illuminating paper [1966] on Russell's theory of descriptions) that Russell himself didn't use the notation (181) which makes definite descriptions into quantifiers. What he did instead was to invent the notation $\imath x\psi$ *and then use it both as a quantifier and as a term*, even though this makes for a contorted syntax. Kaplan detects in this 'a lingering ambivalence' in the mind of the noble lord.

The *third* group of noun phrases express things which can't be said with first-order formulas. Peirce [1885] invented the *two-thirds* quantifier which enables us to say 'At least $\frac{2}{3}$ of the company have white neckties'. (His example.) Peirce's quantifier was unrestricted. It seems more natural, and changes nothing in principle, if we allow a relativisation predicate and write $\frac{2}{3}x(\psi, \phi)$ to mean 'At least $\frac{2}{3}$ of the things x which satisfy ψ satisfy ϕ'.

Can this quantifier be defined away in the spirit of (176)–(179)? Unfortunately not. Let me prove this. By a *functional* I shall mean an expression which is a first-order formula except that formula metavariables may occur in it, and it has no constant symbols except perhaps =. By substituting actual formulas for the metavariables, we get a first-order formula. Two functionals will be reckoned *logically equivalent* if whenever the same formulas are substituted for the metavariables in both functionals, the resulting first-order formulas are logically equivalent. For example the expression $\exists_{\geq 2}x\phi$, viz.

(186) $\exists y(\phi[y/x] \wedge \exists x(\neg x = y \wedge \phi))$,

is a functional which is logically equivalent to $\exists_{\geq 3}x\phi \vee \exists_{=2}x\phi$. Notice that we allow the functional to change some variables which it binds, so as to avoid clash of variables.

A theorem of Skolem [1919] and Behmann [1922] (cf. [Ackermann, 1962, pp. 41–47]) states that *if a functional binds only one variable in each in-*

serted formula, then it is logically equivalent to a combination by $\neg, \wedge$ *and* $\vee$ *of equations* $y = z$ *and functionals of the form* $\exists_{=n} x\chi$ *where* χ *is a functional without quantifiers.* Suppose now that we could define away the quantifier $\frac{2}{3}x(,)$. The result would be a functional binding just the variable x in ψ and ϕ, so by the Skolem–Behmann theorem we could rewrite it as a propositional compound of a finite number of functionals of the form $\exists_{=n} x\chi$, and some equations. (The equations we can forget, because the meaning of $\frac{2}{3}x(\psi, \phi)$ shows that it has no significant free variables beyond those in ψ or ϕ.) If n is the greatest integer for which $\exists_{=n} x$ occurs in the functional, then the functional is incapable of distinguishing any two numbers greater than n, so that it can't possibly express that one of them is at least $\frac{2}{3}$ times the other.

A harder example is

(187) The average Briton speaks at least two-thirds of a foreign language.

I take this to mean that if we add up the number of foreign languages spoken by each Briton, and divide the sum total by the number of Britons, then the answer is at least $\frac{2}{3}$. Putting $\psi(x)$ for 'x is a Briton' and $\phi(x, y)$ for 'y is a foreign language spoken by x', this can be symbolised as $\{Av\frac{2}{3}xy\}(\psi, \phi)$. Can the quantifier $\{Av\frac{2}{3}xy\}$ be defined away in a first-order language? Again the answer is no. This time the Skolem–Behmann result won't apply directly, because $\{Av\frac{2}{3}xy\}$ binds two variables, x and y, in the second formula ϕ. But indirectly the same argument will work. $\frac{2}{3}x(\psi, \phi)$ expresses just the same thing as $\forall z(\psi[z/x] \to \{Av\frac{2}{3}xy\}(\psi, z = x \wedge \phi[y/x] \wedge \psi[y/x]))$. Hence if $\{Av\frac{2}{3}xy\}$ could be defined away, then so could $\frac{2}{3}x$, and we have seen that this is impossible.

Barwise and Cooper [1981] made a thorough study of the logical properties of natural language noun phrases. See also [Montague, 1970; Montague, 1973], particularly his discussion of 'the'. Van Benthem and Doets (this Volume) have a fuller discussion of things not expressible in first-order language.

III: The Expressive Power of First-order Logic

22 AFTER ALL THAT, WHAT IS FIRST-ORDER LOGIC?

It may seem perverse to write twenty-one sections of a chapter about elementary (i.e. first-order) logic without ever saying what elementary logic is. But the easiest definition is ostensive: elementary logic is the logic that we have been doing in Sections 1–18 above. But then, why set *that* logic apart from any other? What particular virtues and vices does it have?

At first sight the Traditional Logician might well prefer a stronger logic. After all, the more valid argument schemas you can find him the happier he is. But in fact Traditional Logicians tend to draw a line between what is 'genuinely logic' and what is really mathematics. The 'genuine logic' usually turns out to be a version of first-order logic.

One argument often put forward for this choice of 'genuine logic' runs along the following lines. In English we can group the parts of speech into two groups. The first group consists of *open classes* such as nouns, verbs, adjectives. These classes expand and contract as people absorb new technology or abandon old-fashioned morality. Every word in these classes carries its own meaning and subject-matter. In the second group are the *closed classes* such as pronouns and conjunctions. Each of these classes contains a fixed, small stock of words; these words have no subject-matter, and their meaning lies in the way they combine with open-class words to form phrases. Quirk and Greenbaum [1973, p.18] list the following examples of closed-class words: the, a, that, this, he, they, anybody, one, which, of, at, in, without, in spite of, and, that, when, although, oh, ah, ugh, phew.

The Traditional Logicians' claim is essentially this: 'genuine logic' is the logic which assembles those valid argument schemas in which open-class words are replaced by schematic letters and closed-class words are not. Quirk and Greenbaum's list already gives us $\wedge$ 'and', $\neg$ 'without', $\forall$ 'anybody', $\exists$ 'a', and of course the words 'not', 'if', 'then', 'or' are also closed-class words. The presence of 'at', 'in spite of' and 'phew' in their list doesn't imply we ought to have added any such items to our logic, because these words don't play any distinctive role in arguments. (The presence of 'when' is suggestive though.) Arguably it is impossible to express second-order conditions in English without using open-class words such as 'set' or 'concept'.

It's a pretty theory. Related ideas run through Quine's [1970]. But for myself I can't see why features of the surface grammar of a few languages that we know and love should be considered relevant to the question what is 'genuine logic'.

We turn to the Proof Theorist. His views are not very helpful to us here. As we saw in Section 20 above, there is in principle no difference between a first-order proof calculus and a non-first-order one. Still, he is likely to make the following comment, which is worth passing on. For certain kinds of application of logic in mathematics, a stronger logic may lead to weaker results. To quote one example among thousands: in a famous paper [1965] Ax and Kochen showed that for each positive integer d there are only finitely many primes which contradict a conjecture of Artin about d. Their proof used heavy set theory and gave no indication what these primes were. Then Cohen [1969] found a proof of the same result using no set-theoretic assumptions at all. From his proof one can calculate, for each d, what the bad primes are. By using the heavy guns, Ax and Kochen had gained intuition but lost information. The moral is that we should think twice before strengthening our logic. The mere fact that a thing is provable in a weaker logic may lead us to further information.

We turn to the Model Theorist. He was probably taught that 'first-order' means we only quantify over elements, not over subsets of the domain of a structure. By now he will have learned (Section 21 above) that some kinds of quantification over elements are not first-order either.

What really matters to a Model Theorist in his language is the interplay of strength and weakness. Suppose he finds a language which is so weak that it can't tell a Montagu from a Capulet. Then at once he will try to use it to prove things about Capulets, as follows. First he shows that something is true for all Montagus, and then he shows that this thing is expressible in his weak language L. Then this thing must be true for at least one Capulet too, otherwise he could use it to distinguish Montagus from Capulets in L. If L is bad enough at telling Montagus and Capulets apart, he may even be able to deduce that *all* Capulets have the feature in question. These methods, which are variously known as *overspill* or *transfer* methods, can be extremely useful if Montagus are easier to study than Capulets.

It happens that first-order languages are excellent for encoding finite combinatorial information (e.g. about finite sequences or syntax), but hopelessly bad at distinguishing one infinite cardinal or infinite ordering from another infinite cardinal or infinite ordering. This particular combination makes first-order model theory very rich in transfer arguments. For example the whole of Abraham Robinson's non-standard analysis [Robinson, 1967] is one vast transfer argument. The Model Theorist will not lightly give up a language which is as splendidly weak as the Upward and Downward Löwenheim–Skolem Theorem and the Compactness Theorem (Section 17 above) show first-order languages to be.

This is the setting into which Per Lindström's theorem came (Section 27 below). He showed that any language which has as much coding power as first-order languages, but also the same weaknesses which have just been mentioned, must actually be a first-order language in the sense that each of its sentences has exactly the same models as some first-order sentence.

23 SET THEORY

In 1922 Skolem described a set of first-order sentences which have become accepted, with slight variations, as the definitive axiomatisation of set theory and hence in some sense a foundation for mathematics. Skolem's axioms were in fact a first-order version of the informal axioms which Zermelo [1908] had given, together with one extra axiom (Replacement) which Fraenkel [1922] had also seen was necessary. The axioms are known as ZFC—Zermelo–Fraenkel set theory with Choice. They are listed in Appendix C below and developed in detail in [Suppes, 1972] and [Levy, 1979].

When these axioms are used as a foundation for set theory or any other part of mathematics, they are read as being about a particular collection V, the class of all sets. Mathematicians differ about whether we have any access to this collection V independently of the axioms. Some writers [Gödel, 1947] believe V is the standard model of the axioms, while others [von Neumann, 1925] regard the symbol 'V' as having no literal meaning at all. But everybody agrees that the axioms have a standard reading, namely as being about V. In this the axioms of ZFC differ from, say, the axioms for group theory, which are never read as being about The Group, but simply as being true in any group.

These axioms form a foundation for mathematics in two different ways. First, some parts of mathematics are directly about sets, so that all their theorems can be phrased quite naturally as statements about V. For example the natural numbers are now often taken to be sets. If they are sets, then the integers, the rationals, the reals, the complex numbers and various vector spaces over the complex numbers are sets too. Thus the whole of real and complex analysis is now recognised as being part of set theory and can be developed from the axioms of ZFC.

Some other parts of mathematics are not about sets, but can be *encoded* in V. We already have an example in Section 17 above, where we converted languages into sets. There are two parts to an encoding. First the entities under discussion are replaced by sets, and we check that all the relations between the original entities go over into relations in V that can be defined within the language of first-order set theory. In the case of our encoded languages, it was enough to note that any finite sequence of sets $a_1, \dots, a_n$ can be coded into an ordered n-tuple $\langle a_1, \dots, a_n \rangle$, and that lengths of sequences, concatenations of sequences and the result of altering one term of a sequence can all be defined. (Cf. [Gandy, 1974].)

The second part of an encoding is to check that all the theorems one wants to prove can be deduced from the axioms of ZFC. Most theorems of elementary syntax can be proved using only the much weaker axioms of Kripke–Platek set theory (cf. [Barwise, 1975]); these axioms plus the axiom of infinity suffice for most elementary model theory too. (Harnik [1985] and [1987] analyses the set-theoretic assumptions needed for various theorems in model theory.) Thus the possibility of encoding pieces of mathematics in set theory rests on two things: first the expressive power of the first-order language for talking about sets, and second the proving power of the set-theoretic axioms. Most of modern mathematics lies within V or can be encoded within it in the way just described. Not all the encodings can be done in a uniform way; see for example Feferman [1969] for a way of handling tricky items from category theory, and the next section below for a trickier item from set theory itself. I think it is fair to say that all of modern mathematics can be encoded in set theory, but it has to be done locally and not all at once, and sometimes there is a perceptible loss of meaning in the encoding. (Incidentally the rival system of *Principia Mathematica*, using a higher-order logic, came nowhere near this goal. As Gödel says of *Principia* in his [1951]: 'it is clear that the theory of real numbers in its present form cannot be obtained'.)

One naturally asks how much of the credit for this universality lies with first-order logic. Might a weaker logic suffice? The question turns out to be not entirely well-posed; if this other logic can in some sense express everything that can be expressed in first-order logic, then in what sense is it 'weaker'? In case any reader feels disposed to look at the question and clarify it, let me mention some reductions to other logics.

First, workers in logic programming or algebraic specification are constantly reducing first-order statements to universal Horn expressions. One can systematise these reductions; see for example Hodges [1993b, Section 10], or Padawitz [1988, Section 4.8]. Second, using very much subtler methods, Tarski and Givant [1987] showed that one can develop set theory within an equational relational calculus $\mathcal{L}^\times$. In their Preface they comment:

> ... $\mathcal{L}^\times$ is equipollent (in a natural sense) to a certain fragment ... of first-order logic having one binary predicate and containing *just three variables.* ... It is therefore quite surprising that $\mathcal{L}^\times$ proves adequate for the formalization of practically all known systems of set theory and hence for the development of all of classical mathematics.

And third, there may be some mileage in the fact that essentially any piece of mathematics can be encoded in an elementary topos (cf. [Johnstone, 1977]).

Amazingly, Skolem's purpose in writing down the axioms of ZFC was to debunk the enterprise: 'But in recent times I have seen to my surprise that so many mathematicians think that these axioms of set theory provide the ideal foundation for mathematics; therefore it seemed to me that the time had come to publish a critique' [Skolem, 1922].

In fact Skolem showed that, since the axioms form a countable first-order theory, they have a countable model $\mathfrak{A}$. In $\mathfrak{A}$ there are 'sets' which satisfy the predicate 'x is uncountable', but since $\mathfrak{A}$ is countable, these 'sets' have only countably many 'members'. This has become known as Skolem's Paradox, though in fact there is no paradox. The set-theoretic predicate 'x is uncountable' is written so as to catch the uncountable elements of V, and there is no reason at all to expect it to distinguish the uncountable elements of other models of set theory. More precisely, this predicate says 'there is no 1–1 function from x to the set ω'. In a model $\mathfrak{A}$ which is different from V, this only expresses that there is no function which is an element of $\mathfrak{A}$ and which is 1–1 from x to ω.

According to several writers the real moral of Skolem's Paradox is that there is no standard model of ZFC, since for any model $\mathfrak{A}$ of ZFC there is another model $\mathfrak{B}$ which is not isomorphic to $\mathfrak{A}$ but is indistinguishable from $\mathfrak{A}$ by first-order sentences. If you have already convinced yourself that the only things we can say about an abstract structure $\mathfrak{A}$ are of the form 'Such-and-such first-order sentences are true in $\mathfrak{A}$', then you should find this argument persuasive. (See [Klenk, 1976; Putnam, 1980] for further discussion.)

Skolem's own explanation of why his argument debunks axiomatic set-theoretic foundations is very obscure. He says in several places that the conclusion is that the meaning of 'uncountable' is relative to the axioms of set theory. I have no idea what this means. The obvious conclusion, surely, is that the meaning of 'uncountable' is relative to the *model*. But Skolem said that he didn't believe in the existence of uncountable sets anyway, and we learn he found it disagreeable to review the articles of people who did [Skolem, 1955].

Contemporary set theorists make free use of non-standard—especially countable—models of ZFC. One usually requires the models to be well-founded, i.e. to have no elements which descend in an infinite sequence

(188) $\cdots \in a_2 \in a_1 \in a_0.$

It is easy to see that this is not a first-order condition on models (for example, Hodges [1972] constructs models of full first-order set theory with arbitrarily long descending sequences of ordinals but no uncountable increasing well-ordered sequences—these models are almost inversely well-founded.) However, if we restrict ourselves to models which are subsets of V, then the statement that such a model contains no sequence (188) can be written as a first-order formula in the language of V. The moral is that it is simply meaningless to classify mathematical statements absolutely as 'first-order' or 'not first-order'. One and the same statement can perfectly well express a second-order condition on structure $\mathfrak{A}$ but a first-order condition on structure $\mathfrak{B}$. (Cf. Section 20 above.)

Meanwhile since the 1950s a number of set theorists have been exploring first-order axioms which imply that the universe of sets is *not* well-founded. Axioms of this kind are called *anti-foundation axioms*; they are rivals to the Foundation (or Regularity) axiom ZF3 in Appendix C below. For many years this work went largely unnoticed, probably because nobody saw any foundational use for it (forgive the pun). But in the 1980s Aczel [1988] saw how to use models of anti-foundation axioms in order to build representations of infinite processes. Barwise generalised Aczel's idea and used non-well-founded sets to represent self-referential phenomena in semantics and elsewhere (cf. [Barwise and Moss, 1996]). Of course there is no problem about describing non-well-founded relations in conventional set theory. The advantage of models of anti-foundation axioms is that they take the membership relation $\in$ itself to be non-well-founded, and it is claimed that this allows us to fall back on other intuitions that we already have about set membership.

24 ENCODING SYNTAX

I begin by showing that the definition of truth in the class V of all sets is not itself expressible in V by a first-order formula. This will demonstrate that there is at least one piece of mathematics which can't be encoded in set theory without serious change of meaning.

As we saw in the previous section, there is no problem about encoding the first-order language L of set theory into V. Without going into details, let me add that we can go one stage further and add to the language L a name for each set; the resulting language L^+ can still be encoded in V as a definable proper class. Let us assume this has been done, so that every formula of L^+ is in fact a set. For each set b, we write $\ulcorner b \urcorner$ for the constant of L^+ which names b. (This is nothing to do with Quine's corners $\ulcorner\ \urcorner$.) When we speak of sentences of L^+ being true in V, we mean that they are true in the structure whose domain is V where '$\in$' is interpreted as set membership and each constant $\ulcorner b \urcorner$ is taken as a name of b.

A class X of sets is said to be *definable* by the formula ψ if for every set α,

(189) $V \models \psi[\alpha/x]$ iff $\alpha \in X$.

Since every set α has a name $\ulcorner \alpha \urcorner$, (189) is equivalent to:

(190) $V \models \psi(\ulcorner \alpha \urcorner/x)$ iff $\alpha \in X$

where I now write $\psi(\ulcorner \alpha \urcorner/x)$ for the result of putting $\ulcorner \alpha \urcorner$ in place of free occurrences of x in ψ.

Suppose now that the class of true sentences of L^+ can be defined by a formula *True* of L^+ with the free variable x. Then for every sentence ϕ of L^+, according to (190),

(191) $V \models$ *True* $(\ulcorner \phi \urcorner/x)$ iff $V \models \phi$.

But since the syntax of L^+ is definable in V, there is a formula χ of L^+ with just x free, such that for every formula ϕ of L^+ with just x free, if $\ulcorner \phi \urcorner = b$ then

(192) $V \models \chi(\ulcorner b \urcorner/x)$ iff $V \models \neg$ *True* $(\ulcorner \phi(\ulcorner b \urcorner/x) \urcorner/x)$.

Now put $b = \ulcorner \chi \urcorner$. Then by (191) and (192),

(193) $V \models \chi(\ulcorner b \urcorner/x)$ iff $V \models$ *True* $(\ulcorner \chi(\ulcorner b \urcorner/x) \urcorner/x)$ iff $V \models \neg\chi(\ulcorner b \urcorner/x)$.

Evidently the two ends of (193) make a contradiction. Hence the class of true sentences of L can't be defined by any formula of L. Thus we have shown that

THEOREM 15. *The class of pairs* $\langle \phi, g \rangle$ *where* ϕ *is a formula of the language* L *of set theory, g is an assignment in* V *and* $V \vDash \phi[g]$, *is not definable in* V *by any formula of the language* L^+ *of set theory with names for arbitrary sets.*

This is one version of Tarski's [1935] *theorem on the undefinability of truth.* Another version, with essentially the same proof, is:

THEOREM 16. *The class of sentences* ϕ *of* L *which are true in* V *is not definable in* V *by any formula of* L.

Of course the set b of all true sentences of L would be definable in V if we allowed ourselves a name for b. Hence the difference between Theorems 15 and 16. These two theorems mean that the matter of truth in V has to be handled either informally or not at all.

Lévy [1965] gives several refined theorems about definability of truth in V. He shows that truth for certain limited classes of sentences of L^+ can be defined in V; in fact each sentence of L^+ lies in one of his classes. As I remarked earlier, everything can be encoded, but not all at once.

Tarski's argument was based on a famous paper of Gödel [1931b], to which I now turn. When formalising the language of arithmetic it is common to include two restricted quantifiers $(\forall x < y)$ and $(\exists x < y)$, meaning respectively 'for all x which are less than y' and 'there is an x which is less than y, such that'. A formula in which every quantifier is restricted is called a Δ_0 formula. Formulas of form $\forall \vec{x} \phi$ and $\exists \vec{x} \phi$, where ϕ is a Δ_0 formula, are said to be Π_1 and Σ_1 respectively. (See under 'Arithmetical hierarchy' in van Dalen (this Volume).)

N shall be the structure whose elements are the natural numbers; each natural number is named by an individual constant $\ulcorner n \urcorner$, and there are relations or functions giving 'plus' and 'times'. A relation on the domain of N which is defined by a Π_1 or Σ_1 formula is said to be a Π_1 or Σ_1 relation respectively. Some relations can be defined in both ways; these are said to be Δ_1 relations. The interest of these classifications lies in a theorem of Kleene [1943].

THEOREM 17. *An n-place relation R on the natural numbers is* Δ_1 *iff there is a computational test which decides whether any given n-tuple is in R; an n-tuple relation R on the natural numbers is* Σ_1 *iff a computer can be programmed to print out all and only the n-tuples in R.*

Hilbert in [1926], the paper that started this whole line of enquiry, had laid great stress on the fact that we can test the truth of a Δ_0 sentence in a finite number of steps, because each time we meet a restricted quantifier we have only to check a finite number of numbers. This is the central idea of the proofs from left to right in Kleene's equivalences. The other directions are proved by encoding computers into N; see Theorems 2.5 and 2.14 in Van Dalen (this Volume).

Now all grammatical properties of a sentence can be checked by mechanical computation. So we can encode the language of first-order Peano arithmetic into N in such a way that all the grammatical notions are expressed by Δ_1 relations. (This follows from Theorem 17, but Gödel [1931b] wrote out an encoding explicitly.) We shall suppose that this has been done, so that from now on every formula or symbol of the language of arithmetic is simply a number. Thus every formula ϕ is a number which is named by the individual constant $\ulcorner\phi\urcorner$. Here $\ulcorner\phi\urcorner$ is also a number, but generally a different number from ϕ; $\ulcorner\phi\urcorner$ is called the *Gödel number* of ϕ. Note that if T is any mechanically describable theory in the language of arithmetic, then a suitably programmed computer can spew out all the consequences of T one by one, so that by Kleene's equivalences (Theorem 17), the set of all sentences ϕ such that $T \vdash \phi$ is a Σ_1 set.

We need one other piece of general theory. Tarski *et al.* [1953] describe a sentence Q in the language of arithmetic which is true in N and has the remarkable property that for every Σ_1 sentence ϕ,

(194) $Q \vdash \phi$ iff $N \vDash \phi$.

We shall use these facts to show that the set of numbers n which are not sentences deducible from Q is not a Σ_1 set. Suppose it were a Σ_1 set, defined by the Σ_1 formula ψ. Then for every number n we would have

(195) $N \vDash \psi(\ulcorner n\urcorner/x) \quad \text{iff} \quad \text{not}(Q \vdash n).$

Now since all syntactic notions are Δ_1, with a little care one can find a Σ_1 formula χ with just x free, such that for every formula ϕ with just x free, if $\ulcorner\phi\urcorner = n$ then

(196) $N \vDash \chi(\ulcorner n\urcorner/x) \quad \text{iff} \quad N \vDash \psi(\ulcorner\phi(\ulcorner n\urcorner/x)\urcorner/x).$

Putting $n = \ulcorner\chi\urcorner$ we get by (194), (195) and (196):

(197)
$$\begin{aligned} N \vDash \chi(\ulcorner n\urcorner/x) \quad &\text{iff} \quad N \vDash \psi(\ulcorner\chi(\ulcorner n\urcorner/x)\urcorner/x) \\ &\text{iff not}(Q \vdash \chi(\ulcorner n\urcorner/x)) \\ &\text{iff not}(N \vDash \chi(\ulcorner n\urcorner/x)) \end{aligned}$$

where the last equivalence is because $\chi(\ulcorner n\urcorner/x)$ is a Σ_1 sentence. The two ends of (197) make a contradiction; so we have proved that the set of numbers n which are not sentences deducible from Q is not Σ_1. Hence the set of numbers which *are* deducible is not Δ_1, and therefore by Theorem 17 there is no mechanical test for what numbers belong to it. We have proved: there is no mechanical test which determines, for any given sentence ϕ of the language of arithmetic, whether or not $\vdash (Q \to \phi)$. This immediately implies Church's theorem [1936]:

THEOREM 18. *There is no mechanical test to determine which sentences of first-order languages are logically valid.*

Now we can very easily prove a weak version of Gödel's [1931b] incompleteness theorem too. Let P be first-order Peano arithmetic. Then it can be shown that $P \vdash Q$. Hence from (194) we can infer that (194) holds with P in place of Q. So the same argument as above shows that the set of non-consequences of P is not Σ_1. If P had as consequences all the sentences true in N, then the non-consequences of P would consist of (i) the sentences ϕ such that $P \vdash \neg\phi$, and (ii) the numbers which are not sentences. But these together form a Σ_1 set. Hence, as Gödel proved,

THEOREM 19. *There are sentences which are true in N but not deducible from P.*

Finally Tarski's theorem (Theorems 15, 16) on the undefinability of truth applies to arithmetic just as well as to set theory. A set of numbers which is definable in N by a first-order formula is said to be *arithmetical.* Tarski's theorem on the undefinability of truth in N states:

THEOREM 20. *The class of first-order sentences which are true in N is not arithmetical.*

Van Benthem and Doets (this Volume) show why Theorem 19 implies that there can be no complete formal proof calculus for second-order logic.

For work connecting Gödel's argument with modal logic, see Boolos [1979; 1993] and Smoryński (Volume 9 of this *Handbook*).

25 SKOLEM FUNCTIONS

When Hilbert interpreted $\exists x\phi$ as saying in effect 'The element x which I choose satisfies ϕ' (cf. Section 15 above), Brouwer accused him of 'causing mathematics to degenerate into a game' [Hilbert, 1928]. Hilbert was delighted with this description, as well he might have been, since games which are closely related to Hilbert's idea have turned out to be an extremely powerful tool for understanding quantifiers.

Before the technicalities, here is an example. Take the sentence

(198) Everybody in Croydon owns a dog.

Imagine a game G: you make the first move by producing someone who lives in Croydon, and I have to reply by producing a dog. I win if and only if the dog I produced belongs to the person you produced. Assuming that I have free access to other people's dogs, (198) is true if and only if I can always win the game G. This can be rephrased: (198) is true if and only if there is a function F assigning a dog to each person living in Croydon, such that whenever we play G, whatever person x you produce, if I retaliate with dog $F(x)$ then I win. A function F with this property is called a *winning strategy* for me in the game G. By translating (198) into a statement about winning strategies, we have turned a statement of form $\forall x \exists y \phi$ into one of form $\exists F \forall x \psi$.

Now come the technicalities. For simplicity, I shall assume that our language L doesn't contain $\bot, \to$ or $\leftrightarrow$, and that all occurrences of $\neg$ are immediately in front of atomic formulas. The arguments of Sections 5 and 15 show that every first-order formula is logically equivalent to one in this form, so the theorems proved below hold without this restriction on L. $\mathfrak{A}$ shall be a fixed L-structure. For each formula ϕ of L and assignment g in $\mathfrak{A}$ to the free variables of ϕ, we shall define a game $G(\mathfrak{A}, \phi; g)$ to be played by two players $\forall$ and $\exists$ (male and female). The definition of $G(\mathfrak{A}, \phi; g)$ is by induction on the complexity of ϕ, and it very closely follows the definition of $\models$ in Section 14:

1. If ϕ is atomic then neither player makes any move in $G(\mathfrak{A}, \phi; g)$ or $G(\mathfrak{A}, \neg\phi; g)$; player $\exists$ wins $G(\mathfrak{A}, \phi; g)$ if $\mathfrak{A} \models \phi[g]$, and she wins $G(\mathfrak{A}, \neg\phi; g)$ if $\mathfrak{A} \models \neg\phi[g]$; player $\forall$ wins iff player $\exists$ doesn't win.

2. Suppose ϕ is $\psi \wedge \chi$, and g_1 and g_2 are respectively the restrictions of g to the free variables of ψ, χ; then player $\forall$ has the first move in $G(\mathfrak{A}, \phi; g)$, and the move consists of deciding whether the game shall proceed as $G(\mathfrak{A}, \psi; g_1)$ or as $G(\mathfrak{A}, \chi; g_2)$.

3. Suppose ϕ is $\psi \vee \chi$, and g_1, g_2 are as in (2); then player $\exists$ moves by deciding whether the game shall continue as $G(\mathfrak{A}, \psi; g_1)$ or $G(\mathfrak{A}, \chi; g_2)$.

4. If ϕ is $\forall x \psi$ then player $\forall$ chooses an element α of $\mathfrak{A}$, and the game proceeds as $G(\mathfrak{A}, \psi; g, \alpha/x)$.

5. If ϕ is $\exists x \psi$ then player $\exists$ chooses an element α of $\mathfrak{A}$, and the game proceeds as $G(\mathfrak{A}, \psi; g, \alpha/x)$.

If g is an assignment suitable for ϕ, and h is the restriction of g to the free variables of ϕ, then $G(\mathfrak{A}, \phi; g)$ shall be $G(\mathfrak{A}, \phi; h)$. When ϕ is a sentence, h is empty and we write the game simply as $G(\mathfrak{A}, \phi)$.

The quantifier clauses for these games were introduced in [Henkin, 1961]. It is then clear how to handle the other clauses; see [Hintikka, 1973, Chapter V]. Lorenzen [1961; 1962] (cf. also Lorenzen and Schwemmer [1975]) described similar games, but in his versions the winning player had to *prove* a sentence, so that his games turned out to define intuitionistic provability where ours will define truth. (Cf. Felscher (Volume 7 of this *Handbook*.) In Lorenzen [1962] one sees a clear link with cut-free sequent proofs.

A *strategy* for a player in a game is a set of rules that tell him how he should play, in terms of the previous moves of the other player. The strategy is called *winning* if the player wins every time he uses it, regardless of how the other player moves. Leaving aside the game-theoretic setting, the next result probably ought to be credited to Skolem [1920]:

THEOREM 21. *Assume the axiom of choice* (cf. Appendix C). *Then for every* L*-structure* $\mathfrak{A}$*, every formula* ϕ *of* L *and every assignment* g *in* $\mathfrak{A}$ *which is suitable for* $\phi, \mathfrak{A} \vDash \phi[g]$ *iff player* $\exists$ *has a winning strategy for the game* $G(\mathfrak{A}, \phi; g)$.

Theorem 21 is proved by induction on the complexity of ϕ. I consider only clause (4), which is the one that needs the axiom of choice. The 'if' direction is not hard to prove. For the 'only if', suppose that $\mathfrak{A} \vDash \forall x\psi[g]$, where g is an assignment to the free variables of $\forall x\psi$. Then $\mathfrak{A} \vDash \psi[g, \alpha/x]$ for every element α; so by the induction assumption, player $\exists$ has a winning strategy for each $G(\mathfrak{A}, \psi; g, \alpha/x)$. Now *choose* a winning strategy S_α for player $\exists$ in each game $G(\mathfrak{A}, \psi; g, \alpha/x)$. Player $\exists$'s winning strategy for $G(\mathfrak{A}, \phi; g)$ shall be as follows: wait to see what element α player $\forall$ chooses, and then follow S_α for the rest of the game.

Theorem 21 has a wide range of consequences. First, it shows that games can be used to give a definition of truth in structures. In fact this was Henkin's purpose in introducing them. See Chapter III of Hintikka [1973] for some phenomenological reflections on this kind of truth-definition.

For the next applications we should bear in mind that *every first-order formula can be converted into a logically equivalent first-order formula which is prenex, i.e. with all its quantifiers at the left-hand end.* (Cf. (127).) When ϕ is prenex, a strategy for player $\exists$ takes a particularly simple form. It consists of a set of functions, one for each existential quantifier in ϕ, which tell player $\exists$ what element to choose, depending on what elements were chosen by player $\forall$ at earlier universal quantifiers.

For example if ϕ is $\forall x\exists y\forall z\exists tR(x, y, z, t)$, then a strategy for player $\exists$ in $G(\mathfrak{A}, \phi)$ will consist of two functions, a 1-place function F_y and a 2-place function F_t. This strategy will be winning if and only if

(199) for all elements α and γ, $\mathfrak{A} \vDash R(x, y, z, t)[\alpha/x, F_y(\alpha)/y, \gamma/z, F_t(\alpha, \gamma)/t]$.

Statement (199) can be paraphrased as follows. Introduce new function symbols f_y and f_t. Write $\phi\hat{}$ for the sentence got from ϕ by removing the existential quantifiers and then putting $f_y(x), f_t(x, z)$ in place of y, t respectively. So $\phi\hat{}$ is $\forall x\forall zR(x, f_y(x), z, f_t(x, z))$. We expand $\mathfrak{A}$ to a structure $\mathfrak{A}\hat{}$ by adding interpretations $I_{\mathfrak{A}\hat{}}(f_y)$ and $I_{\mathfrak{A}\hat{}}(f_t)$ for the new function symbols; let F_y and F_t be these interpretations. Then by (199),

(200) F_y, F_t are a winning strategy for player $\exists$ in $G(\mathfrak{A}, \phi)$ iff $\mathfrak{A}\hat{} \vDash \phi\hat{}$.

Functions F_y, F_t which do satisfy either side of (200) are called *Skolem functions for* ϕ. Putting together (200) and Theorem 21, we get

(201) $\mathfrak{A} \models \phi$ iff by adding functions to $\mathfrak{A}$ we can get a structure $\mathfrak{A}^\wedge$ such that $\mathfrak{A}^\wedge \models \phi^\wedge$.

A sentence $\phi^\wedge$ can be defined in the same way whenever ϕ is any prenex sentence; (201) will still apply. Note that $\phi^\wedge$ is of the form $\forall\vec{x}\psi$ where ψ has no quantifiers; a formula of this form is said to be *universal.*

From (201) we can deduce:

THEOREM 22. *Every prenex first-order sentence* ϕ *is logically equivalent to a second-order sentence* $\exists\vec{f}\phi^\wedge$ *in which* $\phi^\wedge$ *is universal.*

In other words, we can always push existential quantifiers to the left of universal quantifiers, provided that we convert the existential quantifiers into second-order function quantifiers $\exists\vec{f}$. Another consequence of (201) is:

LEMMA 23. *For every prenex first-order sentence* ϕ *we can effectively find a universal sentence* $\phi^\wedge$ *which has a model iff* ϕ *has a model.*

Because of Lemma 23, $\phi^\wedge$ is known as the *Skolem normal form of* ϕ *for satisfiability.*

Lemma 23 is handy for simplifying various logical problems. But it would be handier still if no function symbols were involved. At the end of Section 18 we saw that anything that can be said with a function constant can also be said with a relation constant. However, in order to make the implication from right to left in (201) still hold when relations are used instead of functions, we have to require that the relations really do represent functions, in other words some sentences of form (146) must hold. These sentences are $\forall\exists$ sentences, i.e. they have form $\forall\vec{x}\exists\vec{y}\psi$ where ψ has no quantifiers. The upshot is that for every prenex first-order sentence ϕ *without function symbols* we can effectively find an $\forall\exists$ first-order sentence $\phi_\wedge$ *without function symbols but with extra relation symbols*, such that ϕ has a model if and only if $\phi_\wedge$ has a model. The sentence $\phi_\wedge$ is also known as the *Skolem normal form of* ϕ *for satisfiability.*

For more on Skolem normal forms see [Kreisel and Krivine, 1967, Chapter 2].

Skolem also applied Theorem 21 to prove his part of the Löwenheim–Skolem Theorem 14. We say that L-structures $\mathfrak{A}$ and $\mathfrak{B}$ are *elementarily equivalent* to each other if exactly the same sentences of L are true in $\mathfrak{A}$ as in $\mathfrak{B}$. Skolem showed:

THEOREM 24. *If* L *is a language with at most countably many formulas and* $\mathfrak{A}$ *is an infinite* L*-structure, then by choosing countably many elements of* $\mathfrak{A}$ *and throwing out the rest, we can get a countable* L*-structure* $\mathfrak{B}$ *which is elementarily equivalent to* $\mathfrak{A}$.

This is proved as follows. There are countably many sentences of L which are true in $\mathfrak{A}$. For each of these sentences ϕ, player $\exists$ has a winning strategy S_ϕ for $G(\mathfrak{A}, \phi)$. All we need to do is find a countable set X of elements of $\mathfrak{A}$ such that if player $\forall$ chooses his elements from X, all the strategies S_ϕ tell player $\exists$ to pick elements which are in X too. Then X will serve as the domain of $\mathfrak{B}$, and player $\exists$ will win each $G(\mathfrak{B}, \phi)$ by playing the same strategy S_ϕ as for $G(\mathfrak{A}, \phi)$. Starting from any countable set X_0 of elements of $\mathfrak{A}$, let X_{n+1} be X_n together with all elements called forth by any of the strategies S_ϕ when player $\forall$ chooses from X_n; then X can be the set of all elements which occur in X_n for at least one natural number n.

In his paper [1920], Skolem noticed that the proof of Theorem 21 gives us information in a rather broader setting too. Let $\mathfrak{L}_{\omega_1\omega}$ be the logic we get if, starting from first-order logic, we allow formulas to contain conjunctions or disjunctions of countably many formulas at a time. For example, in $\mathfrak{L}_{\omega_1\omega}$ there is an infinite sentence

(202) $\forall x(x = 0 \lor x = 1 \lor x = 2 \lor \cdots)$

which says 'Every element is a natural number'. If we add (202) to the axioms of first-order Peano arithmetic we get a theory whose only models are the natural number system and other structures which are exact copies of it. This implies that the Compactness Theorem (Theorem 13) and the Upward Löwenheim–Skolem Theorem (Theorem 14) both fail when we replace first-order logic by $\mathfrak{L}_{\omega_1\omega}$.

Skolem noticed that the proof of Theorem 21 tells us:

THEOREM 25. *If ϕ is a sentence of the logic $\mathfrak{L}_{\omega_1\omega}$ and $\mathfrak{A}$ is a model of ϕ, then by choosing at most countably many elements of $\mathfrak{A}$ we can get an at most countable structure $\mathfrak{B}$ which is also a model of ϕ.*

So a form of the Downward Löwenheim–Skolem Theorem (cf. Theorem 14) does hold in $\mathfrak{L}_{\omega_1\omega}$.

To return for a moment to the games at the beginning of this section: Hintikka [1996] has pointed out that there is an unspoken assumption that each player is allowed to know the previous choices of the other player. (If I don't know what person in Croydon you have produced, how can I know which dog to choose?) He has proposed that we should recast first-order logic so that this assumption need no longer hold. For example, in his notation, if ϕ is the sentence

(203) $\forall x(\exists y/\forall x)x = y$

then in the game $G(\mathfrak{A}, \phi)$, player $\forall$ chooses an element a of $\mathfrak{A}$, then player $\exists$ chooses an element b of $\mathfrak{A}$ *without being told what a is.* Player $\exists$ wins if and only if $a = b$. (One easily sees that if $\mathfrak{A}$ has at least two elements, then neither player has a winning strategy for this game.) These added slash quantifiers greatly add to the expressive power of first-order logic. For example there is now a sentence which is true in a structure $\mathfrak{A}$ if and only if $\mathfrak{A}$ has infinitely many elements; there is no such sentence of ordinary first-order logic. As a result, the compactness theorem fails for Hintikka's logic, and hence in turn the logic has no complete proof calculus. One can construct a Tarski-style semantics for the new logic (by a slight adaptation of [Hodges, 1997b]), but it has some very odd features. It no longer makes sense to talk of an element *satisfying* a formula; instead one has to use the notion of a set of elements *uniformly satisfying* the formula, where 'uniform' means essentially that player $\exists$ doesn't need any forbidden information about which element within the set has been chosen. Hintikka claims, boldly, that the extended logic is in several ways more natural than the usual first-order logic.

26 BACK-AND-FORTH EQUIVALENCE

In this section and the next, we shall prove that certain things are definable by first-order formulas. The original versions of the theorems we prove go back to the mid 1950s. But for us their interest lies in the proofs which Per Lindström gave in [1969]. He very cleverly used the facts (1) that first-order logic is good for encoding finite sequences, and (2) that first-order logic is bad for distinguishing infinite cardinals. His proofs showed that anything we can say using a logic which shares features (1) and (2) with first-order logic can also be said with a first-order sentence; so first-order logic is essentially the only logic with these features.

I should say what we mean by a logic. A *logic* $\mathfrak{L}$ is a family of languages, one for each similarity type, together with a definition of what it is for a sentence of a language L of $\mathfrak{L}$ to be true in an L-structure. Just as in first-order logic, an L-structure is a structure which has named relations and elements corresponding to the similarity type of L. We shall always assume that the analogue of Theorem 1 holds for $\mathfrak{L}$, i.e., that the truth-value of a sentence ϕ in a structure $\mathfrak{A}$ doesn't depend on how $\mathfrak{A}$ interprets constants which don't occur in ϕ.

We shall say that a logic $\mathfrak{L}$ is an *extension of first-order logic* if, roughly speaking, it can do everything that first-order logic can do and maybe a bit more. More precisely, it must satisfy three conditions. (i) Every first-order formula must be a formula of $\mathfrak{L}$. (ii) If ϕ and ψ are formulas of $\mathfrak{L}$ then so are $\neg\phi, \phi \wedge \psi, \phi \vee \psi, \phi \to \psi, \phi \leftrightarrow \psi, \forall x\phi, \exists x\phi$; we assume the symbols $\neg$ etc. keep their usual meanings. (iii) $\mathfrak{L}$ is *closed under relativisation*. This means that for every sentence ϕ of $\mathfrak{L}$ and every 1-place predicate constant P not in ϕ, there is a sentence $\phi^{(P)}$ such that a structure $\mathfrak{A}$ is a model of $\phi^{(P)}$ if and only if the part of $\mathfrak{A}$ with domain $I_{\mathfrak{A}}(P)$ satisfies ϕ. For example, if $\mathfrak{L}$ can say 'Two-thirds of the elements satisfy $R(x)$', then it must also be able to say 'Two-thirds of the elements which satisfy $P(x)$ satisfy $R(x)$'. First-order logic itself is closed under relativisation; although I haven't called attention to it earlier, it is a device which is constantly used in applications.

The logic $\mathfrak{L}_{\omega_1\omega}$ mentioned in the previous section is a logic in the sense defined above, and it is an extension of first-order logic. Another logic which extends first-order logic is $\mathfrak{L}_{\infty\omega}$; this is like first-order logic except that we are allowed to form conjunctions and disjunctions of arbitrary sets of formulas, never mind how large. Russell's logic, got by adding definite description operators to first-order logic, is another extension of first-order logic though it never enables us to say anything new.

We shall always require logics to obey one more condition, which needs some definitions. L-structures $\mathfrak{A}$ and $\mathfrak{B}$ are said to be *isomorphic* to each other if there is a function F from the domain of $\mathfrak{A}$ to the domain of $\mathfrak{B}$ which is bijective, and such that for all elements $\alpha_0, \alpha_1, \ldots,$ of $\mathfrak{A}$ and every atomic formula ϕ of L,

(204) $\mathfrak{A} \vDash \phi[\alpha_0/x_0, \alpha_1/x_1, \ldots]$ iff $\mathfrak{B} \vDash \phi[F(\alpha_0)/x_0, F(\alpha_1)/x_1, \ldots]$.

It will be helpful in this section and the next if we omit the x_i's when writing conditions like (204); so (205) means the same as (204) but is briefer:

(205) $\mathfrak{A} \vDash \phi[\alpha_0, \alpha_1, \ldots]$ iff $\mathfrak{B} \vDash \phi[F(\alpha_0), F(\alpha_1), \ldots]$.

If (204) or equivalently (205) holds, where F is a bijection from the domain of $\mathfrak{A}$ to that of $\mathfrak{B}$, we say that F is an *isomorphism* from $\mathfrak{A}$ to $\mathfrak{B}$. Intuitively, $\mathfrak{A}$ is isomorphic to $\mathfrak{B}$ when $\mathfrak{B}$ is a perfect copy of $\mathfrak{A}$.

If $\mathfrak{L}$ is a logic, we say that structures $\mathfrak{A}$ and $\mathfrak{B}$ are $\mathfrak{L}$-*equivalent* to each other if every sentence of $\mathfrak{L}$ which is true in one is true in the other. Thus 'elementarily equivalent' means $\mathfrak{L}$-equivalent where $\mathfrak{L}$ is first-order logic. The further condition we impose on logics is this: structures which are isomorphic to each other must also be $\mathfrak{L}$-equivalent to each other. Obviously this is a reasonable requirement. Any logic you think of will meet it.

Now we shall introduce another kind of game. This one is used for comparing two structures. Let $\mathfrak{A}$ and $\mathfrak{B}$ be L-structures. The game $\mathrm{EF}_\omega(\mathfrak{A};\mathfrak{B})$ is played by two players $\forall$ and $\exists$ as follows. There are infinitely many moves. At the ith move, player $\forall$ chooses one of $\mathfrak{A}$ and $\mathfrak{B}$ and then selects an element of the structure he has chosen; then player $\exists$ must pick an element from the other structure. The elements chosen from $\mathfrak{A}$ and $\mathfrak{B}$ at the ith move are written α_i and β_i respectively. Player $\exists$ wins the game if and only if for every atomic formula ϕ of L,

(206) $\mathfrak{A} \vDash \phi[\alpha_0, \alpha_1, \ldots]$ iff $\mathfrak{B} \vDash \phi[\beta_0, \beta_1, \ldots]$.

We say that $\mathfrak{A}$ and $\mathfrak{B}$ are *back-and-forth equivalent* to each other if player $\exists$ has a winning strategy for this game.

The game $\mathrm{EF}_\omega(\mathfrak{A};\mathfrak{B})$ is known as the *Ehrenfeucht–Fraïssé* game of length ω, for reasons that will appear in the next section. One feels that the more similar $\mathfrak{A}$ and $\mathfrak{B}$ are, the easier it ought to be for player $\exists$ to win the game. The rest of this section is devoted to turning this feeling into theorems. For an easy start:

THEOREM 26. *If $\mathfrak{A}$ is isomorphic to $\mathfrak{B}$ then $\mathfrak{A}$ is back-and-forth equivalent to $\mathfrak{B}$.*

Given an isomorphism F from $\mathfrak{A}$ to $\mathfrak{B}$, player $\exists$ should always choose so that for each natural number $i, \beta_i = F(\alpha_i)$. Then she wins. Warning: we are talking set theory now, so F may not be describable in terms which any human player could use, even if he could last out the game.

As a partial converse to Theorem 26:

THEOREM 27. *If $\mathfrak{A}$ is back-and-forth equivalent to $\mathfrak{B}$ and both $\mathfrak{A}$ and $\mathfrak{B}$ have at most countably many elements, then $\mathfrak{A}$ is isomorphic to $\mathfrak{B}$.*

For this, imagine that player $\forall$ chooses his moves so that he picks each element of $\mathfrak{A}$ or $\mathfrak{B}$ at least once during the game; he can do this if both structures are countable. Let player $\exists$ use her winning strategy. When all the α_i's and β_i's have been picked, define F by putting $F(\alpha_i) = \beta_i$ for each i. (The definition is possible because (206) holds for each atomic formula '$x_i = x_j$'.) Comparing (205) with (206), we see that F is an isomorphism. The idea of this proof was first stated by Huntington [1904] and Hausdorff [1914, p. 99] in proofs of a theorem of Cantor about dense linear orderings. Fraïssé [1954] noticed that the argument works just as well for structures as for orderings.

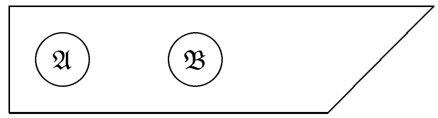
Now we are going to show that whether or not $\mathfrak{A}$ and $\mathfrak{B}$ have countably many elements, if $\mathfrak{A}$ and $\mathfrak{B}$ are back-and-forth equivalent then they are elementarily equivalent. This was known to Fraïssé [1955], and Karp [1965] gave a direct proof of the stronger result that $\mathfrak{A}$ is back-and-forth equivalent to $\mathfrak{B}$ if and only if $\mathfrak{A}$ is $\mathfrak{L}_{\infty\omega}$-equivalent to $\mathfrak{B}$. The interest of our proof (which was extracted from Lindström [1969] by Barwise [1974]) is that it works for any extension of first-order logic which obeys the Downward Löwenheim–Skolem Theorem. To be precise:

THEOREM 28. *Suppose $\mathfrak{L}$ is an extension of first-order logic, and every structure of at most countable similarity type is $\mathfrak{L}$-equivalent to a structure with at most countably many elements. Suppose also that every sentence of $\mathfrak{L}$ has at most countably many distinct symbols. Then any two structures which are back-and-forth equivalent are $\mathfrak{L}$-equivalent to each other.*

Theorem 28 can be used to prove Karp's result too, by a piece of set-theoretic strong-arm tactics called 'collapsing cardinals' (as in [Barwise, 1973]). By Skolem's observation (Theorem 25), Theorem 28 applies almost directly to $\mathfrak{L}_{\omega_1\omega}$ (though one still has to use 'countable fragments' of $\mathfrak{L}_{\omega_1\omega}$—I omit details).

Let me sketch the proof of Theorem 28. Assume all the assumptions of Theorem 28, and let $\mathfrak{A}$ and $\mathfrak{B}$ be L-structures which are back-and-forth equivalent. We have to show that $\mathfrak{A}$ and $\mathfrak{B}$ are $\mathfrak{L}$-equivalent. Replacing $\mathfrak{B}$ by an isomorphic copy if necessary, we can assume that $\mathfrak{A}$ and $\mathfrak{B}$ have no elements in common. Now we construct a jumbo structure:

(207) $\mathfrak{C}$:



The language of $\mathfrak{C}$ shall contain two 1-place predicate constants $\partial^{\mathfrak{A}}$ and $\partial^{\mathfrak{B}}$. Also for each predicate constant R and individual constant c of L the language of $\mathfrak{C}$ shall contain two symbols $R^{\mathfrak{A}}, R^{\mathfrak{B}}$ and $c^{\mathfrak{A}}, c^{\mathfrak{B}}$. The elements in $I_{\mathfrak{C}}(\partial^{\mathfrak{A}})$ are precisely the elements of $\mathfrak{A}$, and each $I_{\mathfrak{C}}(R^{\mathfrak{A}})$ and $I_{\mathfrak{C}}(c^{\mathfrak{A}})$ is to be identical with $I_{\mathfrak{A}}(R)$ and $I_{\mathfrak{A}}(c)$ respectively. Thus $\mathfrak{C}$ contains an exact copy of $\mathfrak{A}$. Likewise with $\mathfrak{B}$ in place of $\mathfrak{A}$. The remaining pieces of $\mathfrak{C}$ outside $\mathfrak{A}$ and $\mathfrak{B}$ consist of enough set-theoretic apparatus to code up all finite sequences of elements of $\mathfrak{A}$ and $\mathfrak{B}$. Finally the language of $\mathfrak{C}$ shall have a 2-place predicate constant S which encodes the winning strategy of player $\exists$ in the game $\mathrm{EF}_{\omega}(\mathfrak{A};\mathfrak{B})$ as follows:

(208) $I_{\mathfrak{C}}(S)$ contains exactly those ordered pairs $\langle\langle\gamma_0,\ldots,\gamma_{n-1}\rangle,\gamma_n\rangle$ such that γ_n is the element which player $\exists$'s winning strategy tells her to play if player $\forall$'s previous moves were $\gamma_0,\ldots,\gamma_{n-1}$.

Now we wish to show that any sentence $\mathfrak{L}$ which is true in $\mathfrak{A}$ is true also in $\mathfrak{B}$, and *vice versa*. Since each sentence of $\mathfrak{L}$ contains at most countably many symbols, we can assume without any loss of generality that the similarity type of $\mathfrak{A}$ and $\mathfrak{B}$ has just countably many symbols; hence the same is true for $\mathfrak{C}$, and thus by the assumption in Theorem 28, $\mathfrak{C}$ is $\mathfrak{L}$-equivalent to a structure $\mathfrak{C}'$ with at most countably many elements. The sets $I_{\mathfrak{C}'}(\partial^{\mathfrak{A}})$ and $I_{\mathfrak{C}'}(\partial^{\mathfrak{B}})$ of $\mathfrak{C}'$ define L-structures $\mathfrak{A}'$ and $\mathfrak{B}'$ which are $\mathfrak{L}$-equivalent to $\mathfrak{A}$ and $\mathfrak{B}$ respectively, since everything we say in $\mathfrak{L}$ about $\mathfrak{A}$ can be rewritten as a statement about $\mathfrak{C}$ using $\partial^{\mathfrak{A}}$ and the $R^{\mathfrak{A}}$ and $c^{\mathfrak{A}}$. (Here we use the fact that $\mathfrak{L}$ allows relativisation.)

Since $\mathfrak{L}$ contains all first-order logic, everything that we can say in a first-order language about $\mathfrak{C}$ must also be true in $\mathfrak{C}'$. For example we can say in first-order sentences that for every finite sequence $\gamma_0,\ldots,\gamma_{n-1}$ of elements of $\mathfrak{A}$ or $\mathfrak{B}$ there is a unique element γ_n such that $\langle\langle\gamma_0,\ldots,\gamma_{n-1}\rangle,\gamma_n\rangle$ is in $I_{\mathfrak{C}}(S)$; also that if player $\exists$ in $\mathrm{EF}_{\omega}(\mathfrak{A};\mathfrak{B})$ reads $I_{\mathfrak{C}}(S)$ as a strategy for her, then she wins. So all these things must be true also for $\mathfrak{A}', \mathfrak{B}'$ and $I_{\mathfrak{C}'}(S)$. (The reader can profitably check for himself that all this can be coded into first-order sentences, but if he gets stuck he can consult [Barwise, 1974] or [Flum, 1975].)

Therefore $\mathfrak{A}'$ is back-and-forth equivalent to $\mathfrak{B}'$. But both $\mathfrak{A}'$ and $\mathfrak{B}'$ are bits of $\mathfrak{C}'$, so they have at most countably many elements. Hence by Theorem 27, $\mathfrak{A}'$ is isomorphic to $\mathfrak{B}'$ and therefore $\mathfrak{A}'$ is $\mathfrak{L}$-equivalent to $\mathfrak{B}'$.

But $\mathfrak{A}'$ was $\mathfrak{L}$-equivalent to $\mathfrak{A}$ and $\mathfrak{B}'$ was $\mathfrak{L}$-equivalent to $\mathfrak{B}$. So finally we deduce that $\mathfrak{A}$ and $\mathfrak{B}$ are $\mathfrak{L}$-equivalent.

In our definition of logics, we allowed the formulas to include some items that go beyond first-order logic, but we made no change in the class of L-structures. The methods of this section, and many of those of the next section too (in particular Theorem 29), still work if one restricts attention to finite structures. Ebbinghaus and Flum [1995] explore the implications of this fact, with an eye on complexity theory.

27 LINDSTRÖM'S THEOREM

Theorem 28 showed that any extension of first-order logic which obeys a form of the Downward Löwenheim–Skolem Theorem is in a sense no stronger than the infinitary logic $\mathfrak{L}_{\infty\omega}$. This result is relatively shallow and not terribly useful; the logic $\mathfrak{L}_{\infty\omega}$ is quite powerful and not very well understood. (See Van Benthem and Doets [this Volume].) Lindström [1969] found a stronger and more subtle result: he showed that if in addition $\mathfrak{L}$ obeys a form of the Compactness Theorem or the Upward Löwenheim–Skolem Theorem then every sentence of $\mathfrak{L}$ has exactly the same models as some first-order sentence. Since a first-order sentence contains only finitely many symbols, this result evidently needs some finiteness restriction on the sentences of $\mathfrak{L}$. So from now on we shall assume that *all similarity types are finite and have no function symbols.*

Lindström's argument relies on some detailed information about Ehrenfeucht–Fraïssé games. The Ehrenfeucht–Fraïssé game $\mathrm{EF}_n(\mathfrak{A};\mathfrak{B})$ of length n, where n is a natural number, is fought and won exactly like $\mathrm{EF}_\omega(\mathfrak{A};\mathfrak{B})$ except that the players stop after n moves. We say that the structures $\mathfrak{A}$ and $\mathfrak{B}$ are *n-equivalent* if player $\exists$ has a winning strategy for the game $\mathrm{EF}_n(\mathfrak{A};\mathfrak{B})$. If $\mathfrak{A}$ and $\mathfrak{B}$ are back-and-forth equivalent then they are n-equivalent for all n; the converse is not true.

Ehrenfeucht–Fraïssé games of finite length were invented by Ehrenfeucht [1960] as a means of showing that two structures are elementarily equivalent. He showed that if two structures $\mathfrak{A}$ and $\mathfrak{B}$ are n-equivalent for all finite n then $\mathfrak{A}$ and $\mathfrak{B}$ are elementarily equivalent (which follows easily from Theorem 28), and that if the similarity type is finite and contains no function symbols, then the converse holds too. Fraïssé's definitions were different, but in his [1955] he proved close analogues of Ehrenfeucht's theorems, including an analogue of the following:

THEOREM 29. *Let* L *be a first-order language. Then for every natural number n there is a finite set of sentences $\sigma_{n,1},\ldots,\sigma_{n,j_n}$ of* L *such that:*

1. *every* L*-structure $\mathfrak{A}$ is a model of exactly one of $\sigma_{n,1},\ldots,\sigma_{n,j_n}$; if $\mathfrak{A}\vDash\sigma_{n,i}$ we say that $\mathfrak{A}$ has n-type $\sigma_{n,i}$;*

2. L-*structures* $\mathfrak{A}$ *and* $\mathfrak{B}$ *are* n*-equivalent iff they have the same* n*-type.*

Theorem 29 is best proved by defining a more complicated game. Suppose $\gamma_0, \ldots, \gamma_{k-1}$ are elements of $\mathfrak{A}$ and $\delta_0, \ldots, \delta_{k-1}$ are elements of $\mathfrak{B}$. Then the game $\mathrm{EF}_n(\mathfrak{A}, \gamma_0, \ldots, \gamma_{k-1}; \mathfrak{B}, \delta_0, \ldots, \delta_{k-1})$ shall be played exactly like $\mathrm{EF}_n(\mathfrak{A}; \mathfrak{B})$, but at the end when elements $\alpha_0, \ldots, \alpha_{n-1}$ of $\mathfrak{A}$ and $\beta_0, \ldots, \beta_{n-1}$ of $\mathfrak{B}$ have been chosen, player $\exists$ wins if and only if for every atomic formula ϕ,

$$(209) \quad \mathfrak{A} \vDash \phi[\gamma_0, \ldots, \gamma_{k-1}, \alpha_0, \ldots, \alpha_{n-1}] \\ \text{iff } \mathfrak{B} \vDash \phi[\delta_0, \ldots, \delta_{k-1}, \beta_0, \ldots, \beta_{n-1}].$$

So this game is harder for player $\exists$ to win than $\mathrm{EF}_n(\mathfrak{A}; \mathfrak{B})$ was. We say that $\langle \mathfrak{A}, \gamma_0, \ldots, \gamma_{k-1} \rangle$ is n*–equivalent* to $\langle \mathfrak{B}, \delta_0, \ldots, \delta_{k-1} \rangle$ if player $\exists$ has a winning strategy for the game $\mathrm{EF}_n(\mathfrak{A}, \gamma_0, \ldots, \gamma_{k-1}; \mathfrak{B}, \delta_0, \ldots, \delta_{k-1})$. We assert that for each finite k and n there is a finite set of formulas $\sigma^k_{n,1}, \sigma^k_{n,2}$ etc. of L such that

1. for every L-structure $\mathfrak{A}$ and elements $\gamma_0, \ldots, \gamma_{k-1}$ of $\mathfrak{A}$ there is a unique i such that $\mathfrak{A} \vDash \sigma^k_{n,i}[\gamma_0, \ldots, \gamma_{k-1}]$; this $\sigma^k_{n,i}$ is called the n*-type* of $\langle \mathfrak{A}, \gamma_0, \ldots, \gamma_{k-1} \rangle$;
2. $\langle \mathfrak{A}, \gamma_0, \ldots, \gamma_{k-1} \rangle$ and $\langle \mathfrak{B}, \delta_0, \ldots, \delta_{k-1} \rangle$ are n-equivalent iff they have the same n-type.

Theorem 29 will then follow by taking k to be 0. We prove the assertion above for each k by induction on n.

When $n = 0$, for each k there are just finitely many sequences $\langle \mathfrak{A}, \gamma_0, \ldots, \gamma_{k-1} \rangle$ which can be distinguished by atomic formulas. (Here we use the fact that the similarity type is finite and there are no function symbols.) So we can write down finitely many formulas $\sigma^k_{0,1}, \sigma^k_{0,2}$ etc. which distinguish all the sequences that can be distinguished.

When the formulas have been constructed and (1), (2) proved for the number n, we construct and prove them for $n+1$ as follows. Player $\exists$ has a winning strategy for $\mathrm{EF}_{n+1}(\mathfrak{A}, \gamma_0, \ldots, \gamma_{k-1}; \mathfrak{B}, \delta_0, \ldots, \delta_{k-1})$ if and only if she can make her first move so that she has a winning strategy from that point onwards, i.e. if she can ensure that α_0 and β_0 are picked so that

$$\langle \mathfrak{A}, \gamma_0, \ldots, \gamma_{k-1}, \alpha_0 \rangle \text{ is } n\text{-equivalent to } \langle \mathfrak{B}, \delta_0, \ldots, \delta_{k-1}, \beta_0 \rangle.$$

In other words, using (2) for n which we assume has already been proved, player $\exists$ has this winning strategy if and only if for every element α of $\mathfrak{A}$ there is an element β of $\mathfrak{B}$ so that

$$\langle \mathfrak{A}, \gamma_0, \ldots, \gamma_{k-1}, \alpha \rangle \text{ has the same } n\text{-type as } \langle \mathfrak{B}, \delta_0, \ldots, \delta_{k-1}, \beta \rangle,$$

and *vice versa* with $\mathfrak{A}$ and $\mathfrak{B}$ reversed. But this is equivalent to the condition:

$$\text{for every } i,\quad \mathfrak{A} \vDash \exists x_k \sigma^{k+1}_{n,i}[\gamma_0, \ldots, \gamma_{k-1}] \text{ iff } \mathfrak{B} \vDash \exists x_k \sigma^{k+1}_{n,i}[\delta_0, \ldots, \delta_{k-1}].$$

It follows that we can build suitable formulas $\sigma^k_{n+1,i}$ by taking conjunctions of formulas of form $\exists x_k \sigma^{k+1}_{n,i}$ or $\neg\exists x_k \sigma^{k+1}_{n,i}$, running through all the possibilities.

When the formulas $\sigma^k_{n,i}$ have all been defined, we take $\sigma_{n,i}$ to be $\sigma^0_{n,i}$. Thus Theorem 29 is proved.

Barwise [1975, Chapter VII.6] describes the formulas $\sigma^k_{n,i}$ in detail in a rather more general setting. The sentences $\sigma_{n,i}$ were first described by Hintikka [1953] (cf. also [Hintikka, 1973, Chapter XI]), but their meaning was mysterious until Ehrenfeucht's paper appeared. We shall call the sentences *Hintikka sentences*. Hintikka proved that every first-order sentence is logically equivalent to a (finite) disjunction of Hintikka sentences. We shall prove this too, but by Lindström's proof [1969] which assumes only some general facts about the expressive power of first-order logic; so the proof will show that any sentence in any logic with this expressive power has the same models as some first-order sentence, viz. a disjunction of Hintikka sentences. Lindström proved:

THEOREM 30. *Let $\mathfrak{L}$ be any extension of first-order logic with the two properties:*

(a) (Downward Löwenheim–Skolem) If a sentence ϕ of $\mathfrak{L}$ has an infinite model then ϕ has a model with at most countably many elements.

(b) Either (Upward Löwenheim–Skolem) if a sentence of $\mathfrak{L}$ has an infinite model then it has one with uncountably many elements; or (Compactness) if Δ is a theory in $\mathfrak{L}$ such that every finite set of sentences from Δ has a model then Δ has a model.

Then every sentence of $\mathfrak{L}$ has exactly the same models as some first-order sentence.

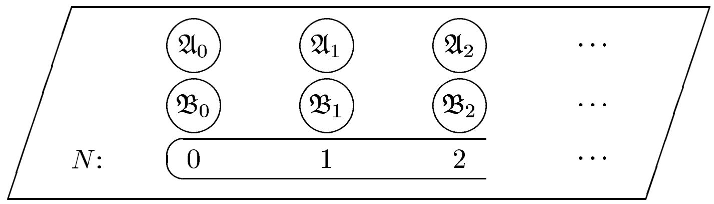
The proof is by the same kind of coding as the proof of Theorem 28. Instead of proving Theorem 30 directly, we shall show:

THEOREM 31. *Let $\mathfrak{L}$ be any extension of first-order logic obeying (a) and (b) as in Theorem 30, and let ϕ and ψ be sentences of $\mathfrak{L}$ such that no model of ϕ is also a model of ψ. Then for some integer n there is a disjunction σ of Hintikka sentences $\sigma_{n,i}$ such that $\phi \vDash \sigma$ and $\psi \vDash \neg\sigma$.*

To get Theorem 30 from Theorem 31, let ψ be $\neg\phi$.

Suppose then that Theorem 31 is false. This means that there exist sentences ϕ and ψ of $\mathfrak{L}$ with no models in common, and for every natural number n there is no disjunction of Hintikka sentences $\sigma_{n,i}$ which separates the models of ϕ from the models of ψ. So by Theorem 29 there are, for each n, n-equivalent structures $\mathfrak{A}_n$ and $\mathfrak{B}_n$ such that $\mathfrak{A}_n$ is a model of ϕ and $\mathfrak{B}_n$ is a model of ψ. By (a) we can assume that $\mathfrak{A}_n$ and $\mathfrak{B}_n$ have at most countably many elements (since the sentences $\sigma_{n,i} \wedge \phi$ and $\sigma_{n,i} \wedge \psi$ are both in $\mathfrak{L}$).

So now once again we build a mammoth model $\mathfrak{C}$:



The coding is more complicated this time. $\mathfrak{C}$ contains a copy of the natural numbers N, picked out by a predicate constant ∂^N. There are 2-place predicate constants $\partial^{\mathfrak{A}}, \partial^{\mathfrak{B}}$. $I_{\mathfrak{C}}(\partial^{\mathfrak{A}})$ contains just those pairs $\langle \alpha, n\rangle$ such that n is a natural number and α is an element of $\mathfrak{A}_n$. Similarly with the $\mathfrak{B}_n$. Also $\mathfrak{C}$ has constants which describe each $\mathfrak{A}_n$ and $\mathfrak{B}_n$ completely, and $\mathfrak{C}$ contains all finite sequences of elements taken from any $\mathfrak{A}_n$ or $\mathfrak{B}_n$, together with enough set theory to describe lengths of sequences etc. There is a relation $I_{\mathfrak{C}}(S)$ which encodes the winning strategies for player $\exists$ in all games $\mathrm{EF}_n(\mathfrak{A}_n, \mathfrak{B}_n)$. Finally $\mathfrak{C}$ can be assumed to have just countably many elements, so we can incorporate a relation which sets up a bijection between N and the whole of the domain of $\mathfrak{C}$.

We shall need the fact that everything salient about $\mathfrak{C}$ can be said in *one single sentence* χ of $\mathfrak{L}$. Since N is in $\mathfrak{C}$ and we can build in as much set-theoretic equipment as we please, this is no problem, bearing in mind that $\mathfrak{L}$ is an extension of first-order logic. Barwise [1974] and Flum [1975] give details.

Now by (b), the sentence χ has a model $\mathfrak{C}'$ in which some 'infinite' number ∞ comes after all the 'natural numbers' $I_{\mathfrak{C}'}(0), I_{\mathfrak{C}'}(1), I_{\mathfrak{C}'}(2), \ldots$ in $I_{\mathfrak{C}'}(\partial^N)$. If the Upward Löwenheim–Skolem property holds, then this is because the N-part of any uncountable model of χ must have the same cardinality as the whole model, in view of the bijection which we incorporated. If on the other hand the Compactness property holds, we follow the construction of non-standard models in Section 20 above.

By means of $I_{\mathfrak{C}'}(\partial^{\mathfrak{A}})$ and $I_{\mathfrak{C}'}(\partial^{\mathfrak{B}})$, the structure $\mathfrak{C}'$ encodes structures $\mathfrak{A}'_\infty$ and $\mathfrak{B}'_\infty$, and $I_{\mathfrak{C}'}(S)$ encodes a winning strategy for player $\exists$ in the game $\mathrm{EF}_\infty(\mathfrak{A}'_\infty; \mathfrak{B}'_\infty)$. All this is implied by a suitable choice of χ. The game $\mathrm{EF}_\infty(\mathfrak{A}'_\infty; \mathfrak{B}'_\infty)$ turns out to be bizarre and quite unplayable; but the important point is that if player $\exists$ has a winning strategy for this game, then she has one for the shorter and entirely playable game $\mathrm{EF}_\omega(\mathfrak{A}'_\infty; \mathfrak{B}'_\infty)$. Hence $\mathfrak{A}'_\infty$ and $\mathfrak{B}'_\infty$ are back-and-forth equivalent.

But now χ records that all the structures encoded by $\partial^{\mathfrak{A}}$ are models of ϕ, while those encoded by $\partial^{\mathfrak{B}}$ are models of ψ. Hence $\mathfrak{A}'_\infty \vDash \phi$ but $\mathfrak{B}'_\infty \vDash \psi$. Since ϕ and ψ have no models in common, it follows that $\mathfrak{B}'_\infty \vDash \neg\phi$. The final step is to use assumption (a), the Downward Löwenheim–Skolem property, to prove a slightly sharpened version of Theorem 28. To be precise, since $\mathfrak{A}'_\infty$ and $\mathfrak{B}'_\infty$ are back-and-forth equivalent and $\mathfrak{A}'_\infty$ is a model of the sentence ϕ of $\mathfrak{L}, \mathfrak{B}'_\infty$ must also be a model of ϕ. (The proof is like that in Section 26, but we use the fact that the similarity type is finite and has no function symbols in order to boil down the essential properties of $\mathfrak{C}$ into a single sentence.) So we have reached a contradiction, and Theorem 31 is proved.

The proof of Theorem 31, less the last paragraph, adapts to give a proof of *Craig's Interpolation Lemma* for predicate logic:

LEMMA 32. *Let ϕ and ψ be sentences of first-order predicate logic such that $\phi \vDash \neg\psi$. Then there is a first-order sentence σ such that $\phi \vDash \sigma, \psi \vDash \neg\sigma$, and every constant symbol which occurs in σ occurs both in ϕ and ψ.*

Let $\mathfrak{L}$ in the proof of Theorem 31 be first-order logic and let L be the first-order language whose constants are those which occur both in ϕ and in ψ. Using Section 18, we can assume that L has no function symbols. If $\mathfrak{A}$ is any model of ϕ, then we get an L-structure $\mathfrak{A}|\mathrm{L}$ by discarding all constant symbols not in L, without changing the elements or the interpretations of the symbols which are in L. Likewise for every model $\mathfrak{B}$ of ψ. Now suppose that the conclusion of Lemma 32 fails. Then for each natural number n there is no disjunction σ of Hintikka sentences $\sigma_{n,i}$ *in the language* L such that $\phi \vDash \sigma$ and $\psi \vDash \neg\sigma$, and hence there are models $\mathfrak{A}_n, \mathfrak{B}_n$ of ϕ, ψ respectively, such that $\mathfrak{A}_n|\mathrm{L}$ is n-equivalent to $\mathfrak{B}_n|\mathrm{L}$. Proceed now as in the proof of Theorem 31, using the Compactness and Downward Löwenheim–Skolem Theorems to find a countable $\mathfrak{C}'$ with an infinite natural number ∞. Excavate models $\mathfrak{A}'_\infty, \mathfrak{B}'_\infty$ of ϕ, ψ from $\mathfrak{C}'$ as before, noting this time that $\mathfrak{A}'_\infty|\mathrm{L}$ is back-and-forth equivalent to $\mathfrak{B}'_\infty|\mathrm{L}$. Then by Theorem 27, since $\mathfrak{A}'_\infty|\mathrm{L}$ and $\mathfrak{B}'_\infty|\mathrm{L}$ are countable and back-and-forth equivalent, they are isomorphic. It follows that we can add to $\mathfrak{A}'_\infty$ interpretations of those symbols which are in ψ but not in L, using $\mathfrak{B}'_\infty$ as a template. Let $\mathfrak{D}$ be the resulting structure. Then $\mathfrak{D} \vDash \phi$ since $\mathfrak{A}'_\infty \vDash \phi$, and $\mathfrak{D} \vDash \psi$ since $\mathfrak{B}'_\infty \vDash \psi$. This contradicts the assumption that $\phi \vDash \neg\psi$. Hence Lemma 32 is proved.

Craig himself [1957b] used his interpolation result to give a proof of *Beth's Definability Theorem* [Beth, 1953]:

THEOREM 33. *Let* L *be a first-order language and Δ a first-order theory which uses the language* L *together with one extra n-place predicate constant R. Suppose that for every* L*-structure $\mathfrak{A}$ there is at most one way of adding to $\mathfrak{A}$ an interpretation of R so that the resulting structure is a model of Δ.*

Then Δ *has a consequence of form* $\forall x_1, \ldots, x_n(R(x_1, \ldots, x_n) \leftrightarrow \phi)$, *where* ϕ *is a formula in the language* L.

Time's wingèd chariot prevents a proper discussion of implicit and explicit definitions here, but Beth's theorem is proved in Section 5.5 of [Hodges, 1997a], and Section 2.2 of Chang and Keisler [1973]. There is some useful background on implicit definitions in [Suppes, 1957, Chapter 8]. Craig's and Beth's results have interested philosophers of science; see e.g. [Sneed, 1971].

28 LAWS OF THOUGHT?

This section is new in the second edition. I am not sure that it belongs at Section 28, but this was the simplest place to add it.

Frege fought many battles against the enemies of sound reason. One battle which engaged some of his best energies was that against *psychologism*. Psychologism, put briefly, was the view that the proper definitions of logical notions (such as validity) make essential reference to the contents of minds. Today psychologism in first-order logic is a dead duck; not necessarily because Frege convinced anybody, but simply because there is no room for any mention of minds in the agreed definitions of the subject. The question whether the sequent

$$p \wedge q \vdash p$$

is valid has nothing more to do with minds than it has to do with the virginity of Artemis or the war in Indonesia.

Still, psychology fights back. The next generation has to learn the subject—and so we find ourselves asking: How does one teach logic? How does one learn it? How far do people think logically anyway, without benefit of logic texts? and what are the mental mechanisms involved?

During the 1980s a number of computer programs for teaching elementary logic came onto the market. Generally they would give the student a sequent and allow him or her to build a formal proof on the screen; then they would check it for correctness. Sometimes they would offer hints on possible ways to find a proof. One can still find such programs today, but mostly they are high-tech practical aids for working computer scientists, and they work in higher-order logic as happily as in first-order. (There is a review of teaching packages in [Goldson, Reeves and Bornat, 1993].) To a great extent the introductory teaching packages were driven out by a better program, *Tarski's World*. This was a sophisticated stand-alone Macintosh program put on the market in 1986 by a team of logicians and computer scientists at Stanford University led by Jon Barwise and John Etchemendy [1991].

Tarski's World teaches the notation of first-order logic, by means of the Hintikka games which we studied in Section 25 above. The student sees on the screen a formal sentence, together with a 'world' which consists of a checker board with various objects on it, some labelled with constant symbols. The predicate symbols in the sentence all have fixed meanings such as 'x is a tetrahedron' or 'x is between y and z'. The student is invited to guess whether the given world makes the sentence true or false, and to defend the guess by playing a game against the machine. (A little later but independently, a group in Omsk produced a similar package for teaching logic to students in Siberia. The Russian version didn't use the notion of games, and its 'worlds' consisted of graphs.)

As it stands, *Tarski's World* is no use for learning about logical consequence: in the first place it contains no proof theory, and in the second place the geometrical interpretations of the predicate symbols are built into the program, so that there is no possibility of constructing counterexamples in general—even small ones. Barwise and Etchemendy found an innovative way to plug the gap. Their next computer package, *Hyperproof* [Barwise and Etchemendy, 1994], consists of a natural deduction theorem prover for first-order logic, together with a device that allows students to represent facts pictorially rather than by sentences. Thus the picture for 'a is a small tetrahedron' is a small tetrahedron labelled a. The picture for 'a is small' is subtler: we have to represent a without showing what shape it is, so the picture is a small paper bag labelled a. There are devices for reading off sentences from pictures, and for adjusting pictures to fit stated sentences. Proofs are allowed to contain both sentences and pictures.

The language is limited to a small number of predicates with fixed meanings: 'x is between y and z', 'x likes y' and a few others. The student is allowed (in fact encouraged) to use geometrical knowledge about the properties of betweenness and the shape of the picture frame. As this suggests, the package aims to teach the students to reason, rather than teaching them logical theory. (On pictorial reasoning in first-order logic, see [Hammer, 1995] and his references.)

There has already been some research on how good *Hyperproof* is at teaching students to reason, compared with more 'syntactic' logic courses. Stenning, Cox and Oberlander [1995] found that one can divide students into two groups—which they call DetHi and DetLo—in terms of their performance on reasoning tests before they take a logic course. DetHi students benefit from *Hyperproof*, whereas a syntactic logic course tends if anything to make them less able to reason about positions of blocks in space. For this spatial reasoning, DetLo students gain more advantage from a syntactic course than from *Hyperproof*. Different patterns emerge on other measures of reasoning skill. Stenning *et al.* comment:

> ...the evidence presented here already indicates both that dif-

> ferent teaching methods can induce opposite effects in different groups of students, and that the *same* teaching method administered in a strictly controlled computerised environment using the same examples, and the same advice can induce different groups of students to develop quite distinct reasoning styles.

We need replications and extensions of this research, not least because there are several ways in which logic courses can differ. *Hyperproof* is more pictorial than any other logic course that I know. But it also belongs with those courses that give equal weight to deduction and consistency, using both proofs and counterexamples; this is a different dimension, and Stenning *et al.* suggest that it might account for some of their findings. Another feature is that students using computer logic programs get immediate feedback from the computer, unlike students learning in a class from a textbook.

These findings are a good peg to hang several other questions on. First, do classes in first-order logic really help students to do anything except first-order logic? Before the days of the Trade Descriptions Act, one early twentieth-century textbook of syllogisms advertised them as a cure for blushing and stammering. (I quote from memory; the book has long since disappeared from libraries.) Psychological experimenters have usually been much more pessimistic, claiming that there is very little transfer of skills from logic courses to any other kind of reasoning. For example Nisbett, Fong, Lehman and Cheng [1987] found that if you want to improve a student's logical skills (as measured by the Wason selection task mentioned below—admittedly a narrow and untypical test), you should teach her two years of law, medicine or psychology; a standard undergraduate course in logic is completely ineffectual. On the other hand Stenning *et al.* [1995] found that a logic course gave an average overall improvement of about 12% on the Analytical Reasoning score in the US Graduate Record Exam (I thank Keith Stenning for this figure). Their results suggest that the improvement may vary sharply with the kind of logic course, the kind of student and the kind of test.

Second, what is the brute native competence in first-order reasoning of a person with average intelligence and education but no specific training in logic? One of the most thorough-going attempts to answer this question is the work of Lance Rips [1994]. Rips writes a theorem-proving program called PSYCOP, which is designed to have more or less the same proficiency in first-order reasoning as the man on the Clapham omnibus. He defends it with a large amount of empirical evidence. A typical example of a piece of reasoning which is beyond PSYCOP is:

> NOT (IF Calvin passes history THEN Calvin will graduate).
> *Therefore* Calvin passes history.

One has to say straight away that the man on the Clapham omnibus has never seen the basic symbols of first-order logic, and there could be a great deal of slippage in the translation between first-order formalism and the words used in the experiments. In Rips' work there certainly is some slippage. For example he regards $\forall x \exists y \neg\phi(x,y)$ as the same sentence as $\neg\exists x \forall y \phi(x,y)$, which makes it impossible for him to ask whether people are successful in deducing one from the other—even though the two forms suggest quite different sentences of English.

It might seem shocking that there are simple first-order inferences which the average person can't make. One suspects that this must be a misdescription of the facts. Anybody who does suspect as much should look at the astonishing 'selection task' experiment of P. C. Wason [1966], who showed that in broad daylight, with no tricks and no race against a clock, average subjects can reliably and repeatedly be brought to make horrendous mistakes of truth-table reasoning. This experiment has generated a huge amount of work, testing various hypotheses about what causes these mistakes; see [Manktelow and Over, 1990].

Third, what are the mental mechanisms that an untrained person uses in making logical deductions? Credit for raising this as an experimental issue goes to P. N. Johnson-Laird, who with his various collaborators has put together a considerable body of empirical facts (summarised in Johnson-Laird and Byrne [1991], see also the critiques in *Behavioral and Brain Sciences*, **16**, 323–380, 1993). Unfortunately it is hard for an outsider to see what thesis Johnson-Laird is aiming to prove with these facts. He uses some of the jargon of logical theory to set up a dichotomy between rule-based reasoning and model-based reasoning, and he claims that his evidence supports the latter against the former. But for anybody who comes to it from the side of logical theory, Johnson-Laird's dichotomy is a nonsense. If it has any meaning at all, it can only be an operational one in terms of the computer simulation which he offers, and I hope the reader can make more sense of that than I could. Perhaps two things emerge clearly. The first is that what he calls model-based reasoning is meta-level—it is reasoning about reasoning; which leaves us asking what his theory of object-level reasoning can be. The second claim to emerge from the mist is that we regularly use a form of proof-by-cases, and the main cause of making deductions that we shouldn't have done is that we fail to list all the necessary cases. This is an interesting suggestion, but I was unable to see how the theory explains the cases where we fail to make deductions that we should have done.

It would be a pity to end on a negative note. This section has shown, I hope, that at the end of the millenium first-order logic is still full of surprises for the old hands and new opportunities for young researchers.

ACKNOWLEDGEMENTS

I thank the many people who commented on the first version of this chapter, and especially Franz Guenthner, Hans Kamp and Dirk van Dalen for their detailed improvements and corrections. I am also very much in debt to Jan and Emilia Mycielski who allowed me to type the chapter in their idyllic house in the foothills of the Rockies.

Finally I thank Keith Stenning for his help with the new Section 28 in the second edition, with the usual caution that he is not to be held responsible for any of the opinions expressed there (except those quoted from a paper of his).

Queen Mary and Westfield College, London.

IV: Appendices

These three appendices will show in outline how one can construct a formal calculus of set theory, which in some sense formalises the whole of mathematics. I have put this material into appendices, first because it is turgid, and second because I should hate to resuscitate the dreadful notion that the business of logicians is to produce all-embracing formal systems.

A. A FORMAL PROOF SYSTEM

We shall define a formal proof system for predicate logic with identity. To cover propositional logic too, the language will have some sentence letters. The calculus is a Hilbert-style system.

First we define the *language* L, by describing its similarity type, its set of terms and its set of formulas (cf. Sections 3 and 13 above).

The *similarity type* of L is made up of the following sentence letters, individual constants, predicate constants and function constants. The *sentence letters* are the expressions p_n, where n is a natural number subscript. The *individual constants* are the expressions c_n, where n is a natural number subscript. The *predicate* constants are the expressions P_n^m, where n is a natural number subscript and m is a positive integer superscript. The *function constants* are the expressions f_n^m, where n is a natural number subscript and m is a positive integer superscript. A predicate or function constant is said to be *m-place* if its superscript is m.

The *terms* of L are defined inductively as follows: (i) Every variable is a term, where the *variables* are the expressions x_n with natural number subscript n. (ii) For each function symbol f_n^m, if $\tau_1, \ldots, \tau_m$ are terms then the expression $f_n^m(\tau_1, \ldots, \tau_m)$ is a term. (iii) Nothing is a term except as required by (i) and (ii).

The *formulas* of L are defined inductively as follows: (i) Every sentence letter is a formula. (ii) The expression $\bot$ is a formula. (iii) For each predicate constant R_n^m, if $\tau_1, \ldots, \tau_m$ are terms then the expression $R_n^m(\tau_1, \ldots, \tau_m)$ is a formula. (iv) If σ and τ are terms then the expression $(\sigma = \tau)$ is a formula. (v) If ϕ and ψ are formulas, then so are the expressions $\neg\phi, (\phi \wedge \psi), (\phi \vee \psi), (\phi \to \psi), (\phi \leftrightarrow \psi)$. (vi) For each variable x_n, if ϕ is a formula then so are the expressions $\forall x_n \phi$ and $\exists x_n \phi$. (vii) Nothing is a formula except as required by (i)–(vi).

A full account would now define two further notions, $FV(\phi)$ (the set of variables with free occurrences in ϕ) and $\phi[\tau_1 \cdots \tau_k / x_{i_i} \cdots x_{i_k}]$ (the formula which results when we simultaneously replace all free occurrences of x_{i_j} in ϕ by τ_j, for each $j, 1 \leq j \leq k$, avoiding clash of variables). Cf. Section 13 above.

Now that L has been defined, formulas occurring below should be read as metalinguistic names for formulas of L. Hence we can make free use of the metalanguage abbreviations in Sections 4 and 13.

Now we define the proof system—let us call it **H**. We do this by describing the axioms, the derivations, and the way in which a sequent is to be read off from a derivation. (Sundholm (see Volume 2) describes an alternative Hilbert-style system **CQC** which is equivalent to **H**.)

The *axioms* of **H** are all formulas of the following forms:

H1. $\phi \to (\psi \to \phi)$

H2. $(\phi \to \psi) \to ((\phi \to (\psi \to \chi)) \to (\phi \to \chi))$

H3. $(\neg\phi \to \psi) \to ((\neg\phi \to \neg\psi) \to \phi)$

H4. $((\phi \to \bot) \to \bot) \to \phi$

H5. $\phi \to (\psi \to \phi \wedge \psi)$

H6. $\phi \wedge \psi \to \phi, \quad \phi \wedge \psi \to \psi$

H7. $\phi \to \phi \vee \psi, \quad \psi \to \phi \vee \psi$

H8. $(\phi \to \chi) \to ((\psi \to \chi) \to (\phi \vee \psi \to \chi))$

H9. $(\phi \to \psi) \to ((\psi \to \phi) \to (\phi \leftrightarrow \psi))$

H10. $(\phi \leftrightarrow \psi) \to (\phi \to \psi), \quad (\phi \leftrightarrow \psi) \to \psi \to \phi)$

H11. $\phi[\tau/x] \to \exists x\phi$

H12. $\forall x\phi \to \phi[\tau/x]$

H13. $x = x$

H14. $x = y \to (\phi \to \phi[y/x])$

A *derivation* (or *formal proof*) in **H** is defined to be a finite sequence

(A.1) $\langle\langle\phi_1, m_1\rangle, \ldots, \langle\phi_n, m_n\rangle\rangle$

such that $n \geq 1$, and for each i ($1 \leq i \leq n$) one of the five following conditions holds:

1. $m_i = 1$ and ϕ_i is an axiom;
2. $m_i = 2$ and ϕ_i is any formula of L;
3. $m_i = 3$ and there are j and k in $\{1, \ldots, i-1\}$ such that ϕ_k is $\phi_j \to \phi_i$;
4. $m_i = 4$ and there is j ($1 \leq j < i$) such that ϕ_j has the form $\psi \to \chi$, x is a variable not free in ψ, and ϕ_i is $\psi \to \forall x\chi$;
5. $m_i = 5$ and there is j ($1 \leq j < i$) such that ϕ_j has the form $\psi \to \chi$, x is a variable not free in χ, and ϕ_i is $\exists x\psi \to \chi$.

Conditions 3–5 are called the *derivation rules* of the calculus. They tell us how we can add new formulas to the end of a derivation. Thus (3) says that if ψ and $\psi \to \chi$ occur in a derivation, then we can add χ at the end; this is the rule of *modus ponens.*

The *premises* of the derivation (A.1) are those formulas ϕ_i such that $m_i = 2$. Its *conclusion* is ϕ_n. We say that ψ is *derivable from* $\chi_1, \ldots, \chi_k$ *in the calculus* **H**, in symbols

(A.2) $\chi_1, \ldots, \chi_n \vdash_{\mathbf{H}} \psi$,

if there exists a derivation whose premises are all among $\chi_1, \ldots, \chi_n$ and whose conclusion is ψ.

Remarks

1. The calculus **H** is sound and strongly complete for propositional and predicate logic with identity. (Cf. Section 7; as in Section 15, this says nothing about provable sequents in which some variables occur free.)
2. In practice most logicians would write the formulas of a derivation as a column or a tree, and they would omit the numbers m_i.
3. To prove the completeness of **H** by either the first or the third method in Section 16, one needs to know for all sentences $\chi_1, \ldots, \chi_n$ and ψ,

   (A.3) if $\chi_1, \ldots, \chi_n \vdash_{\mathbf{H}} \psi$ then $\chi_1, \ldots, \chi_{n-1} \vdash_{\mathbf{H}} \chi_n \to \psi$.

   Statement (A.3) is the *Deduction Theorem* for **H**. It remains true if we allow free variables to occur in the formulas, provided that they occur only in certain ways. See [Kleene, 1952, Sections 21–24] for details.

4. Completeness and soundness tell us that if $\chi_1, \ldots, \chi_n$ and ψ are sentences, then (A.2) holds if and only if $\chi_1, \ldots, \chi_n \vDash \psi$. This gives an intuitive meaning to such sequents. But when $\chi_1, \ldots, \chi_n$ and ψ are allowed to be any formulas of L, then to the best of my knowledge there are no natural necessary and sufficient conditions for (A.2) to hold. So it seems impossible to explain what if anything (A.2) tells us, except by referring to the fine details of the calculus **H**. This is a general feature of Hilbert-style calculi for predicate logic, and I submit that it makes them thoroughly inappropriate for introducing undergraduates to logic.

5. If we are thinking of varying the rules of the calculus, or even if we just want a picture of what the calculus is about, it is helpful to have at least a *necessary* condition for (A.2) to hold. The following supplies one. The *universal closure* of ϕ is $\forall y_1, \ldots, y_n \phi$, where $y_1, \ldots, y_n$ are the free variables of ϕ. Let ϕ_1 be the universal closure of $\chi_1 \wedge \cdots \wedge \chi_n$ and ϕ_2 the universal closure of ψ. Then one can show that

   $$\text{(A.4)} \quad \text{if } \chi_1, \ldots, \chi_n \vdash_{\mathbf{H}} \psi \text{ then } \phi_1 \vDash \phi_2.$$

   The proof of (A.4) is by induction on the lengths of derivations. Statement (A.4) is one way of showing that **H** is sound.

6. The following derivation shows that $\vdash_{\mathbf{H}} \exists x(x = x)$:

   (A.5) $x = x$ (axiom H13)
   $x = x \to \exists x(x = x)$ (axiom H11)
   $\exists x(x = x)$ (from above by modus ponens)

   Statement (A.4) shows the reason, namely:

   $$\text{(A.6)} \quad \forall x(x = x \wedge (x = x \to \exists x(x = x))) \vDash \exists x(x = x).$$

   On any reasonable semantic interpretation (cf. Section 14 above), the left-hand side in (A.6) is true in the empty structure but the right-hand side is false. Suppose now that we want to modify the calculus in order to allow empty structures. Then we must alter the derivation rule which took us from left to right in (A.6), and this is the rule of modus ponens. (Cf. Bencivenga (Volume 7 of this *Handbook*.) It is important to note here that even if (A.4) was a tidy two-way implication, the modus ponens rule would *not* express 'ϕ and $\phi \to \psi$ imply ψ', but rather something of the form '$\forall \vec{x}(\phi \wedge (\phi \to \psi))$ implies $\forall \vec{y} \psi$'. As it is, the meaning of modus ponens in **H** is quite obscure. (Cf. [Kleene, 1952, Section 24].)

B. ARITHMETIC

I begin with naive arithmetic, not formal Peano arithmetic. One needs to have at least an intuitive grasp of naive arithmetic in order to understand what a formal system is. In any case [Peano, 1889] reached his axioms by throwing naive arithmetic into fancy symbols.

Naive arithmetic is adequately summed up by the following five axioms, which come from Dedekind [1888; 1967]. Here and below, 'number' means 'natural number', and I start with 0 (Dedekind's first number was 1).

NA1. 0 is a number.

NA2. For every number n there is a next number after n; this next number is called Sn or the *successor* of n.

NA3. Two different numbers never have the same successor.

NA4. 0 is not the successor of any number.

NA5. (Induction axiom) Let K be any set with the properties (i) 0 is in K, (ii) for every number n in K, Sn is also in K. Then every number is in K.

These axioms miss one vital feature of numbers, viz. their order. So we define $<$ as follows. First we define an *initial segment* to be a set K of numbers such that if a number Sn is in K then n is also in K. We say:

(B.1) $m < n$ iff there is an initial segment which contains m but not n.

The definition (B.1) implies:

(B.2) If $m < Sn$ then either $m < n$ or $m = n$.

For future reference I give a proof. Suppose $m < Sn$ but not $m = n$. Then there is an initial segment K such that m is in K and Sn is not in K. Now there are two cases. *Case 1*: n is not in K. Then by (B.1), $m < n$. *Case 2*: n is in K. Then let M be K with n omitted. Since $m \neq n$, M contains m but not n. Also M is an initial segment; for if Sk is in M but k is not, then by the definition of M we must have $k = n$, which implies that Sn is in M and hence in K; contradiction. So we can use M in (B.1) to show $m < n$.

(B.3) For each number m it is false that $m < 0$.

(B.3) is proved 'by induction on m', using the induction axiom NA5. Proofs of this type are written in a standard style, as follows:

Case 1. $m = 0$. Then $m < 0$ would imply by (B.1) that there was a set containing 0 but not 0, which is impossible.

Case 2. $m = Sk$, assuming it proved when $m = k$. Suppose $Sk < 0$. Then by (B.1) there is an initial segment containing Sk and not 0. Since K is an initial segment containing Sk, k is also in K. So by (B.1) again, K shows that $k < 0$. But the induction hypothesis states that not $k < 0$; contradiction.

This is all one would normally say in the proof. To connect it with NA5, let M be the set of all numbers m such that not $m < 0$. The two cases show exactly what has to be shown, according to NA5, in order to prove that every number is in M.

Here are two more provable facts.

(B.4) The relation $<$ is a linear ordering of the numbers (in the sense of (157)–(159) in Section 19 above).

(B.5) Every non-empty set of numbers has a first element.

Fact (B.5) states that the numbers are *well-ordered*, and it is proved as follows. Let X be any set of numbers without a first element. Let Y be the set of numbers not in X. Then by induction on n we show that every number n is in Y. So X is empty.

Fact (B.5) is one way of justifying *course-of-values induction*. This is a style of argument like the proof of (B.3) above, except that in Case 2, instead of proving the result for Sk assuming it was true for k, we prove it for Sk assuming it was true *for all numbers* $\leq k$. In many theorems about logic, one shows that every formula has some property A by showing (i) that every atomic formula has property A and (ii) that if ϕ is a compound formula whose proper subformulas have A then ϕ has A. Arguments of this type are course-of-values inductions on the complexity of formulas.

In naive arithmetic we can justify two important types of definition. The first is sometimes called *recursive definition* and sometimes *definition by induction*. It is used for defining functions whose domain is the set of natural numbers. To define such a function F recursively, we first say outright what $F(0)$ is, and then we define $F(Sn)$ in terms of $F(n)$. A typical example is the recursive definition of addition:

(B.6) $m + 0 = m, \quad m + Sn = S(m + n).$

Here $F(n)$ is $m + n$; the definition says first that $F(0)$ is m and then that for each number n, $F(Sn)$ is $SF(n)$. To justify such a definition, we have to show that there is exactly one function F which satisfies the stated conditions. To show there is *at most* one such function, we suppose that F and G are two functions which meet the conditions, and we prove by induction on n that for every n, $F(n) = G(n)$; this is easy. To show that there is *at least* one is harder. For this we define an *n-approximation* to be a function whose domain is the set of all numbers $< n$, and which obeys the conditions in the recursive definition for all numbers in its domain. Then we show by induction on n (i) that there is at least one n-approximation, and (ii) that if $m < k < n$, f is a k-approximation and g is an n-approximation, then $f(m) = g(m)$. Then finally we define F explicitly by saying that $F(m)$ is the unique number h such that $f(m) = h$ whenever f is an n-approximation for some number n greater than m.

After defining $+$ by (B.6), we can go on to define $\cdot$ by:

(B.7) $m \cdot 0 = 0, \quad m \cdot Sn = m \cdot n + m.$

The functions definable by a sequence of recursive definitions in this way, using equations and previously defined functions, are called *primitive recursive* functions. Van Dalen [this Volume] discusses them further.

There is a course-of-values recursive definition too: in this we define $F(0)$ outright, and then $F(Sn)$ in terms of values $F(k)$ for numbers $k \leq n$. For example if $F(n)$ is the set of all formulas of complexity n, understood as in Section 3 above, then the definition of $F(n)$ will have to refer to the sets $F(k)$ for all $k < n$. Course-of-values definitions can be justified in the same way as straightforward recursive definitions.

The second important type of definition that can be justified in naive arithmetic is also known as *inductive definition*, though it is quite different from the 'definition by induction' above. Let H be a function and X a set. We say that X is *closed under* H if for every element x of X, if x is in the domain of H then $H(x)$ is also in X. We say that X is the *closure* of Y under H if (i) every element of Y is in X, (ii) X is closed under H, and (iii) if Z is any set which includes Y and is closed under H then Z also includes X. (Briefly, 'X is the smallest set which includes Y and is closed under H'.) Similar definitions apply if we have a family of functions $H_1, \ldots, H_k$ instead of the one function H; also the functions can be n-place functions with $n > 1$.

A set is said to be *inductively defined* if it is defined as being the closure of some specified set Y under some specified functions $H_1, \ldots, H_k$. A typical inductive definition is the definition of the set of terms of a language L. The usual form for such a definition is:

1. Every variable and every individual constant is a term.

2. For each function constant f, if f is n-place and $\tau_1, \ldots, \tau_n$ are terms, then the expression $f(\tau_1, \ldots, \tau_n)$ is a term.

3. Nothing is a term except as required by (1) and (2).

Here we are defining the set X of terms. The so-called *basic* clause (1) describes Y as the set of all variables and all individual constants. The *inductive* clause (2) describes the functions H_i, one for each function constant.

Finally the *extremal* clause (3) says that X is the closure of Y under the H_i. (Many writers omit the extremal clause, because it is rather predictable.)

Frege [1884] may have been the first to argue that inductive definitions need to be justified. He kept asking: How do we know that there *is* a smallest set which includes Y and is closed under H? One possible justification runs as follows. We recursively define $F(n)$, for each positive integer n, to be the set of all sequences $\langle b_1, \ldots, b_n \rangle$ such that b_1 is in Y and for every i $(1 \le i < n)$, b_{i+1} is $H(b_i)$. Then we define X to be the set of all b such that for some number n there is a sequence in $F(n)$ whose last term is b. Clearly Y is included in X, and we can show that X is closed under H. If Z is any set which is closed under H and includes Y, then an induction on the lengths of sequences shows that every element of X is in Z.

Naive arithmetic, as described above, is an axiomatic system but not a formal one. Peano [1889] took the first step towards formalising it, by inventing a good symbolism. But the arguments above use quite an amount of set theory, and Peano made no attempt to write down what he was assuming about sets. Skolem [1923] threw out the set theory and made his assumptions precise, but his system was rather weak. First-order Peano arithmetic, a formalisation of the first-order part of Peano's axioms, was introduced in [Gödel, 1931b].

P, or *first-order Peano Arithmetic*, is the following formal system. The constants of the language are an individual constant $\overline{0}$, a 1-place function symbol S and 2-place functions symbols $+$ and $\bullet$, forming terms of form Sx, $(x + y)$, $(x \bullet y)$. Write $\overline{n}$ as an abbreviation for $S \ldots$ (n times)$\ldots S\overline{0}$; the symbols $\overline{n}$ are called *numerals.* We use a standard proof calculus for first-order logic (e.g. the calculus **H** of Appendix A) together with the following axioms:

P1. $\forall xy(Sx = Sy \to x = y)$

P2. $\forall x \neg (Sx = 0)$

P3. (Axiom schema of induction) All sentences of the form $\forall \vec{z}(\phi[\overline{0}/x] \wedge \forall x(\phi \to \phi[Sx/x]) \to \forall x\phi)$

P4. $\forall x(x + \overline{0} = x)$

P5. $\forall xy(x + Sy = S(x + y))$

P6. $\forall x(x \bullet \overline{0} = \overline{0})$

P7. $\forall xy(x \bullet Sy = (x \bullet y) + x)$

The axioms are read as being just about numbers, so that $\forall x$ is read as 'for all numbers x'. In this way the symbols $\overline{0}$ and S in the language take care of axioms NA1 and NA2 without further ado. Axioms NA3 and NA4 appear as **P**1 and **P**2. Since we can refer only to numbers and not to sets, axiom NA5 has to be recast as a condition on those sets of numbers which are definable by first-order formulas; this accounts for the axiom schema of induction, **P**3.

P4–**P**7 are the recursive definitions of addition and multiplication, cf. (B.6) and (B.7) above. In naive arithmetic there was no need to assume these as axioms, because we could *prove* that there are unique functions meeting these conditions. However, the proof used some set-theoretic notions like 'function defined on the numbers $0, \ldots, n-1$', which can't be expressed in a first-order language using just $\overline{0}$ and S. So we have to put the symbols $+$, $\bullet$ into the language—in particular they occur in formulas in the axiom schema of induction—and we have to assume the definitions **P**4 – **P**7 as axioms.

Gödel showed that with the aid of first-order formulas involving only $\overline{0}, S, +$ and $\bullet$, he could explicitly define a number of other notions. For example

(B.8) $x < y$ iff $\exists z(x + Sz = y)$.

Also by using a clever trick with prime numbers he could encode each finite sequence $\langle m_1, m_2, \ldots \rangle$ of numbers as a single number

(B.9) $2^{m_1+1}.3^{m_2+1}.5^{m_3+1}\ldots.$

and he could express the relation 'x is the yth term of the sequence coded by z' by a first-order formula. But then he could carry out 'in **P**' all the parts of naive arithmetic which use only numbers, finite sequences of numbers, finite sequences of finite sequences of numbers, and so on. This includes the argument which justifies primitive recursive definitions. In fact:

1. *For every recursive definition δ of a number function, using just first-order formulas, there is a formula $\phi(x, y)$ such that in* **P** *we can prove that ϕ defines a function obeying δ. (If δ is primitive recursive then ϕ can be chosen to be Σ_1, cf. Section 24.)*

2. *For every inductive definition of a set, where a formula ψ defines the basic set Y and formulas χ define the functions H in the inductive clause, there is a formula $\phi(x)$ such that we can prove in* **P** *that the numbers satisfying ϕ are those which can be reached in a finite number of steps from Y by H. (If ψ and χ are Σ_1 then ϕ can be chosen to be Σ_1.)*

These two facts state in summary form why the whole of elementary syntax can be formalised within **P**.

There are some things that can be said in the language of **P** but not proved or refuted from the axioms of **P**. For example the statement that **P** itself is consistent (i.e. doesn't yield $\bot$) can be formalised in the language of **P**. In [1931b] Gödel showed that this formalised statement is not deducible from **P**, although we all hope it is true.

There are some other things that can't even be said in the language of **P**. For example we can't say in this language that the set X defined by ϕ in (2) above really is the closure of Y under H, because that would involve us in saying that 'if Z is *any set* which includes Y and is closed under H then Z includes X'. In the first-order language of **P** there is no way of talking about 'all sets of numbers'. For the same reason, many statements about real numbers can't be expressed in the language of **P**—even though some can by clever use of rational approximations.

In second-order arithmetic we can talk about real numbers, because real numbers can be represented as sets of natural numbers. Actually the natural numbers themselves are definable up to isomorphism in second-order logic without special arithmetical axioms. In third-order logic we can talk about sets of real numbers, fourth-order logic can talk about sets of sets of real numbers, and so on. Most of the events that take place in any standard textbook of real analysis can be recorded in, say, fifth-order logic. See Van Benthem and Doets [this Volume] for these higher-order logics.

C. SET THEORY

The efforts of various nineteenth-century mathematicians reduced all the concepts of real and complex number theory to one basic notion: classes. So when Frege, in his *Grundgesetze der Arithmetik I* [1893], attempted a formal system which was to be adequate for all of arithmetic and analysis, the backbone of his system was a theory of classes. One of his assumptions was that for every condition there is a corresponding class, namely the class of all the objects that satisfy the condition. Unfortunately this assumption leads to contradictions, as Russell and Zermelo showed. Frege's approach has now been abandoned.

Today the most commonly adopted theory of classes is Zermelo–Fraenkel set theory, ZF. This theory was propounded by Zermelo [1908] as an informal axiomatic theory. It reached its present shape through contributions from Mirimanoff, Fraenkel, Skolem and von Neumann. (Cf. Fraenkel's historical introduction to [Bernays and Fraenkel, 1958].)

Officially ZF is a set of axioms in a first-order language whose only constant is the 2-place predicate symbol $\in$ ('is a member of'). But all set theorists make free use of symbols introduced by definition.

Let me illustrate how a set theorist introduces new symbols. The axiom of Extensionality says that no two different sets have the same members. The Pair-set axiom says that if x and y are sets then there is at least one set which has just x and y as members. Putting these two axioms together, we infer that there is exactly one set with just x and y as members. Introducing a new symbol, we call this set $\{x, y\}$. There are also some definitions which don't depend on the axioms. For example we say x is *included in* y, or a *subset of* y, if every member of x is a member of y. This prompts the definition

(C1) $x \subseteq y$ iff $\forall t(t \in x \to t \in y)$.

The language with these extra defined symbols is in a sense impure, but it is much easier to read than the pure set language with only $\in$, and one can always paraphrase away the new symbols when necessary. In what follows I shall be relentlessly impure. (On introducing new terms by definition, cf. Section 21 above. Suppes [1972] and Levy [1979] are careful about it.)

The first three axioms of ZF are about what kind of things we choose to count as sets. The axiom of Extensionality says that sets will count as equal when they have the same members:

ZF1. (Extensionality) $\forall xy(x \subseteq y \wedge y \subseteq x \to x = y)$

We think of sets as being built up by assembling their members, starting with the empty or null set 0 which has no members:

ZF2. (Null-set) $\forall t\ t \notin 0$ ($x \notin y$ means $\neg(x \in y)$.)

In a formal calculus which proves $\exists x\ x = x$, the Null-set axiom is derivable from the Separation axiom below and can be omitted. The axiom of Regularity (also known as the axiom of Foundation) expresses—as well as one can express it with a first-order statement—that X will not count as a set unless each of the members of x could be assembled together at an earlier stage than x itself. (So for example there is no 'set' x such that $x \in x$.)

ZF3. (Regularity) $\forall x(x = 0 \vee \exists y(y \in x \wedge \forall z(z \in y \to z \notin x)))$.

The next three axioms state that certain collections can be built up:

ZF4. (Pair-set) $\forall xyt(t \in \{x, y\} \leftrightarrow t = x \vee t = y)$

ZF5. (Union) $\forall xt(t \in \bigcup x \leftrightarrow \exists y(t \in y \wedge y \in x))$

ZF6. (Power-set) $\forall xt(t \in \mathcal{P}x \leftrightarrow t \subseteq x)$.

Axioms ZF3–ZF6 allow some constructions. We write $\{x\}$ for $\{x, x\}, x \cup y$ for $\bigcup\{x, y\}, \{x_1, x_2, x_3\}$ for $\{x_1, x_2\} \cup \{x_3\}, \{x_2, \ldots, x_4\}$ for $\{x_1, x_2, x_3\} \cup \{x_4\}$, and so on. Likewise we can form ordered pairs $\langle x, y\rangle = \{\{x\}, \{x, y\}\}$, ordered triplets $\langle x, y, z\rangle = \langle\langle x, y\rangle, z\rangle$ and so on. Building up from 0 we can form $1 = \{0\}, 2 = \{0, 1\}, 3 = \{0, 1, 2\}$ etc.; the axiom of Regularity implies that $0, 1, 2, \ldots$ are all distinct. We can regard $0, 1, 2, \ldots$ as the natural numbers.

We need to be able to express 'x is a natural number' in the language of set theory, without using informal notions like 'and so on'. It can be done as follows. First, following von Neumann, we define $\mathrm{Ord}(x)$, 'x is an ordinal', by:

(C.2) $\mathrm{Ord}(x)$ iff $\bigcup x \subseteq x \wedge \forall yz(y \in x \wedge z \in x \to y \in z \vee z \in y \vee y = z)$.

This somewhat technical definition implies that the ordinals are linearly ordered by $\in$, and that they are well-ordered (i.e. every non-empty set of them has a least element, cf. (B.5) above). We can prove that the first ordinals are $0, 1, 2, \ldots$. Greek letters α, β, γ are used for ordinals. For every ordinal α there is a first greater ordinal; it is written $\alpha + 1$ and defined as $\alpha \cup \{\alpha\}$. For every set X of ordinals there is a first ordinal β which is greater than or equal to every ordinal in X, viz. $\beta = \bigcup X$. Each ordinal β has just one of the following three forms: either $\beta = 0$, or β is a *successor* (i.e. of form $\alpha + 1$), or β is a *limit* (i.e. of form $\bigcup X$ for a non-empty set X of ordinals which has no greatest member). Now the natural numbers can be defined as follows:

(C.3) x is a natural number iff $\mathrm{Ord}(x) \wedge \forall y(y \in x + 1 \to y = 0 \vee y$ is a successor).

The remaining four axioms, ZF7–ZF10, are needed for talking about infinite sets. Each of them says that sets exist with certain properties. Nothing in ZF1–ZF6 implies that there are any infinite sets. We fill the gap by decreeing that the set ω of all natural numbers exists:

ZF7. (Infinity) $\forall t(t \in \omega \leftrightarrow t$ is a natural number).

The next axiom says that within any given set x we can collect together those members w which satisfy the formula $\phi(\vec{z}, w)$. Here ϕ is allowed to be any first-order formula in the language of set theory, and it can mention other sets $\vec{z}$. Strictly ZF8 is an axiom schema and not a single axiom.

ZF8. (Separation) $\forall \vec{z}xt(t \in \{w \in x | \phi\} \leftrightarrow t \in \{x \wedge \phi[t/w]\}$.

For example this tells us that for any sets x and y there is a set whose members are exactly those members w of x which satisfy the formula $w \in y$; in symbols this set is $\{w \in x | w \in y\}$. So we can introduce a new symbol for this set, and write $x \cap y = \{w \in x | w \in y\}$. Similarly we can define: $\bigcap x = \{w \in \bigcup x | \forall z(z \in x \to w \in z)\}, x \times y = \{t \in \mathcal{PP}(x \cup y) | \exists zw(z \in x \wedge w \in y \wedge t = \langle z, w \rangle)\}, x^2 = x \times x$ and more generally $x^{n+1} = x^n \times x$. An *n-place relation* on the set x is a subset of x^n. We can define 'f is a function from x to y', in symbols $f : x \to y$, by:

(C.4) $f : x \to y$ iff $f \subseteq x \times y \wedge \forall w(w \in x \to \exists z \forall t(t = z \leftrightarrow \langle w, t \rangle \in f))$.

We say f is an *n-place function* from x to y if $f : x^n \to y$. When $f : x \to y$, we call x the *domain* of f, and we can define it in terms of f by: $\text{dom}f = \{w \in \bigcup\bigcup f | \exists z \langle w, z\rangle \in f\}$. We define the *value* of f for *argument* w, in symbols $f(w)$, as $\{t \in \bigcup\bigcup\bigcup f | \exists z(\langle w, z\rangle \in f \wedge t \in z)\}$. A *bijection* (or *one-one correspondence*) *from x to y* is a function f such that $f : x \to y$ and every element z of y is of form $f(w)$ for exactly one w in x. A *sequence of length α* is defined to be a function with domain α.

The system of axioms ZF1–ZF8 is sometimes known as *Zermelo set theory*, or *Z* for short. It is adequate for formalising all of naive arithmetic, not just the finite parts that can be axiomatised in first-order Peano arithmetic. The Separation axiom is needed. For example in the proof of (B.2) we had to know that there is a set M whose members are all the members of K except $n; M$ is $\{w \in K | w \neq n\}$.

First-order languages can be defined formally within Z. For example we can define a *similarity type* for predicate logic to be a set whose members each have one of the following forms: (i) $\langle 1, x\rangle$, (ii) $\langle 2, m, x\rangle$ where m is a positive natural number, (iii) $\langle 3, m, x\rangle$ where m is a positive natural number. The elements of form (i) are called *individual constants*, those of form (ii) are the *m-place predicate constants* and those of form (iii) are the *m-place function constants*. *Variables* can be defined as ordered pairs of form $\langle 4, n\rangle$ where n is a natural number. *Terms* can be defined inductively by: (a) Every variable or individual constant is a term. (b) If f is an m-place function constant and $\tau_1, \ldots, \tau_m$ are terms then $\langle 5, f, \tau_1, \ldots, \tau_m\rangle$ is a term. (c) Nothing is a term except as required by (a) and (b). By similar devices we can define the whole language L of a given similarity type X. L-*structures* can be defined to be ordered pairs $\langle A, I\rangle$ where A is a non-empty set and I is a function with domain X, such that for each individual constant c of $X, I(c) \in A$ (and so on as in Section 14). Likewise we can define $\vDash$ for L-structures.

The two remaining axioms of ZF are needed for various arguments in infinite arithmetic.

In Appendix B we saw how one can define functions with domain the natural numbers, by recursion. We want to be able to do the same in set theory, but with any ordinal as the domain. For example if the language L is not countable, then the proof of completeness in Section 16 above will need to be revised so that we build a chain of theories Δ_i for $i \in \alpha$, where α is some ordinal greater than ω. One can try to justify recursive definitions on ordinals, just as we justified definitions in Appendix B. It turns out that one piece of information is missing. We need to know that if a formula defines a function f whose domain is an ordinal, then f is a set. The following axiom supplies this missing information. It says that if a formula ϕ defines a function with domain a set, then the image of this function is again a set:

ZF9. (Replacement)

$$\forall \vec{z} x(\forall ywt(y \in x \wedge \phi \wedge \phi[w/t] \to t = w) \to$$
$$\exists u \forall t(t \in u \leftrightarrow \exists y(y \in x \wedge \phi))).$$

Like Separation, the Replacement axiom is really an axiom schema.

The final axiom is the axiom of Choice, which is needed for most kinds of counting argument. This axiom can be given in many forms, all equivalent in the sense that any one can be derived from any other using ZF1–ZF9. The form given below, Zermelo's Well-ordering principle, means intuitively that the elements of any set can be checked off one by one against the ordinals, and that the results of this checking can be gathered together into a set.

ZF10. (Well-ordering)
$\forall x \exists f \alpha$ (α is an ordinal and f is a bijection from α to x).

Axiom ZF10 is unlike axioms ZF4–ZF9 in a curious way. These earlier axioms each said that there is a set with just such-and-such members. But ZF10 says that a certain set exists (the function f) without telling us what the members of the set are. So arguments which use the axiom of Choice have to be less explicit than arguments which only use ZF1–ZF9.

Using ZF10, the theory of 'cardinality proceeds as follows. The *cardinality* $|x|$ or $x^{=}$ of a set x is the first ordinal α such that there is a bijection from α to x. Ordinals which are the cardinalities of sets are called *cardinals.* Every cardinal is equal to its own cardinality. Every natural number is a cardinal. A set is said to be *finite* if its cardinality is a natural number. The cardinals which are not natural numbers are said to be *infinite.* The infinite cardinals can be listed in increasing order as $\omega_0, \omega_1, \omega_2, \ldots, ; \omega_0$ is ω. For every ordinal α there is an αth infinite cardinal ω_α, sometimes also written as $\aleph_\alpha$. It can be proved that there is no greatest cardinal, using Cantor's theorem that for every set $x, \mathcal{P}(x)$ has greater cardinality than x.

Let me give an example of a principle equivalent to ZF10. If I is a set and for each $i \in I$ a set A_i is given, then $\Pi_I A_i$ is defined to be the set of all functions $f : I \to \bigcup\{A_i | i \in I\}$ such that for each $j \in I, f(j) \in A_j$. $\Pi_I A_i$ is called the *product* of the sets A_i. Then ZF10 is equivalent to the statement: If the sets A_i in a product are all non-empty then their product is also not empty.

The compactness theorem for propositional logic with any set of sentence letters is not provable from ZF1–ZF9. *A fortiori* neither is the compactness theorem for predicate logic. Logicians have dissected the steps between ZF10 and the compactness theorem, and the following notion is one of the results. (It arose in other parts of mathematics too.)

Let I be any set. Then an *ultrafilter* on I is defined to be a subset D of $\mathcal{P}(I)$ such that (i) if a and $b \in D$ then $a \cap b \in D$, (ii) if $a \in D$ and $a \subseteq b \subseteq I$ then $b \in D$, and (iii) for all subsets a of I, exactly one of I and $I - a$ is in D (where $I - a$ is the set of all elements of I which are not in a). For example if $i \in I$ and $D = \{a \in \mathcal{P}(I) | i \in a\}$ then D is an ultrafilter on I; ultrafilters of this form are called *principal* and they are uninteresting. From ZF1–ZF9 it is not even possible to show that there exist any non-principal ultrafilters at all. But using ZF10 one can prove the following principle:

THEOREM C.5 *Let I be any infinite set. Then there exist an ultrafilter D on I and for each $i \in I$ an element $a_i \in D$, such that for every $j \in I$ the set $\{i \in I | j \in a_i\}$ is finite.*

An ultrafilter D with the property described in Theorem C.5 is said to be *regular*. Regular ultrafilters are always non-principal.

To derive the compactness theorem from Theorem C.5, we need to connect ultrafilters with structures. This is done as follows. For simplicity we can assume that the language L has just one constant symbol, the 2-place predicate constant R. Let D be an ultrafilter on the set I. For each $i \in I$, let $\mathfrak{A}_i$ be an L-structure with domain A_i. Define a relation $\sim$ on $\Pi_I A_i$ by:

(C.6) $f \sim g$ iff $\{i \in I | f(i) = g(i)\} \in D$.

Then since D is an ultrafilter, $\sim$ is an equivalence relation; write $f^\sim$ for the equivalence class containing f. Let B be $\{f^\sim | f \in \Pi_i A_i\}$. Define an L-structure $\mathfrak{B} = \langle B, I_\mathfrak{B} \rangle$ by putting

(C.7) $\langle f^\sim, g^\sim \rangle \in I_\mathfrak{B}(R)$ iff $\{i \in I | \langle f(i), g(i) \rangle \in I_{\mathfrak{A}_i}(R)\} \in D$.

(Using the fact that D is an ultrafilter, this definition makes sense.) Then $\mathfrak{B}$ is called the *ultraproduct* of the $\mathfrak{A}_i$ by D, in symbols $\Pi_D \mathfrak{A}_i$ or D-prod $\mathfrak{A}_i$. By a theorem of Jerzy Łoś, if ϕ is any sentence of the first-order language L, then

(C.8) $\Pi_D \mathfrak{A}_i \vDash \phi$ iff $\{i \in I | \mathfrak{A}_i \vDash \phi\} \in D$.

Using the facts above, we can give another proof of the compactness theorem for predicate logic. Suppose that Δ is a first-order theory and every finite subset of Δ has a model. We have to show that Δ has a model. If Δ itself is finite, there is nothing to prove. So assume now that Δ is infinite, and let I in Theorem C.5 be Δ. Let D and the sets a_ϕ ($\phi \in \Delta$) be as in Theorem C.5. For each $i \in \Delta$, the set $\{\phi | i \in a_\phi\}$ is finite, so by assumption it has a model $\mathfrak{A}_i$. Let $\mathfrak{B}$ be $\Pi_D \mathfrak{A}_i$. For each sentence $\phi \in \Delta, a_\phi \subseteq \{i \in \Delta | \mathfrak{A}_i \vDash \phi\}$, so by (ii) in the definition of an ultrafilter, $\{i \in \Delta | \mathfrak{A}_i \vDash \phi\} \in D$. It follows by Łoś's theorem (C.8) that $\mathfrak{B} \vDash \phi$. Hence Δ has a model, namely $\mathfrak{B}$.

There are full accounts of ultraproducts in Bell and Slomson [1969] and Chang and Keisler [1973]. One principle which often turns up when ultraproducts are around is as follows. Let X be a set of subsets of a set I. We say that X has the *finite intersection property* if for every finite subset $\{a_1, \ldots, a_n\}$ of X, the set $a_1 \cap \cdots \cap a_n$ is not empty. The principle states that *if X has the finite intersection property then there is an ultrafilter D on I such that $X \subseteq D$.* This can be proved quite quickly from ZF10.

Some writers refer to ZF1–ZF9, without the axiom of Choice, as ZF; they write ZFC when Choice is included. There are a number of variants of ZF. For example the set-class theory of Gödel and Bernays (cf. [Mendelson, 1987]) allows one to talk about 'the class of all sets which satisfy the formula ϕ' provided that ϕ has no quantifiers ranging over classes. This extension of ZF is only a notational convenience. It enables one to replace axiom schemas by single axioms, so as to get a system with just finitely many axioms.

Another variant allows elements which are not sets—these elements are called *individuals*. Thus we can talk about the set {Geoffrey Boycott} without having to believe that Geoffrey Boycott is a set. In informal set theory of course one considers such sets all the time. But there seems to be no mathematical advantage in admitting individuals into formal set theory; rather the contrary, we learn nothing new and the proofs are messier. A set is called a *pure set* if its members, its members' members, its members' members' members etc. are all of them sets. In ZF all sets are pure.

BIBLIOGRAPHY

The text of Church [1956] is a reliable and thorough source of information on anything that happened in first-order logic before the mid 1950s. The historical survey by Moore [1980] is also valuable.

[Ackermann, 1962] W. Ackermann. *Solvable Cases of the Decision Problem.* North-Holland, Amsterdam, 1962.

[Aczel, 1988] P. Aczel. *Non-well-founded Sets.* CSLI, Stanford CA, 1988.

[Altham and Tennant, 1975] J. E. J. Altham and N. W. Tennant. Sortal quantification. In E. L. Keenan, editor, *Formal Semantics of Natural Language*, pages 46–58. Cambridge University Press, 1975.

[Anderson and Johnstone Jr., 1962] J. M. Anderson and H. W. Johnstone Jr. *Natural Deduction: The Logical Basis of Axiom Systems.* Wadsworth, Belmont, CA., 1962.

[Ax and Kochen, 1965] J. Ax and S. Kochen. Diophantine problems over local fields: I. *American Journal of Mathematics*, 87:605–630, 1965.

[Barwise, 1973] J. Barwise. Abstract logics and $L_{\infty\omega}$. *Annals Math Logic*, 4:309–340, 1973.

[Barwise, 1974] J. Barwise. Axioms for abstract model theory. *Annals Math Logic*, 7:221–265, 1974.

[Barwise, 1975] J. Barwise. *Admissible Sets and Structures.* Springer, Berlin, 1975.

[Barwise and Cooper, 1981] J. Barwise and R. Cooper. Generalized quantifiers and natural langauges. *Linguistics and Philosophy*, 4:159–219, 1981.

[Barwise and Etchemendy, 1991] J. Barwise and J. Etchemendy. *Tarski's World 3.0.* Cambridge University Press, 1991.

[Barwise and Etchemendy, 1994] J. Barwise and J. Etchemendy. *Hyperproof.* CSLI, Stanford, 1994.

[Barwise and Moss, 1996] J. Barwise and L. Moss. *Vicious Circles.* CSLI, Stanford CA, 1996.

[Behmann, 1922] H. Behmann. Beiträge zur Algebra der Logik, insbesondere zum Entscheidungs-problem. *Math Annalen*, 86:163–229, 1922.

[Bell and Machover, 1977] J. L. Bell and M. Machover. *A Course in Mathematical Logic.* North-Holland, Amsterdam, 1977.

[Bell and Slomson, 1969] J. L. Bell and A. B. Slomson. *Models and Ultraproducts.* North-Holland, Amsterdam, 1969.

[Belnap, 1962] N. D. Belnap. Tonk, plonk and plink. *Analysis*, 22:130–134, 1962. Reprinted in [Strawson, 1967,pp. 132–137].

[Benacerraf and Putnam, 1983] P. Benacerraf and H. Putnam, editors. *Philosophy of Mathematics: Selected Readings.* Cambridge University Press, second edition, 1983.

[Bernays, 1942] P. Bernays. Review of Max Steck, 'Ein unbekannter Brief von Gottlob Frege über Hilberts erste Vorlesung über die Grundlagen der Geometrie'. *Journal of Symbolic Logic*, 7:92 f., 1942.

[Bernays and Fraenkel, 1958] P. Bernays and A. A. Fraenkel. *Axiomatic Set Theory.* North-Holland, Amsterdam, 1958.

[Beth, 1953] E. W. Beth. On Padoa's method in the theory of definition. *Koninklijke Nederlandse Akad. van Wetensch*, 56 (ser. A, Math Sciences):330–339, 1953.

[Beth, 1955] E. W. Beth. Semantic entailment and formal derivability. *Mededelingen der Koninklijke Nederlandse Akad. van Wetensch*, afd letterkunde 18, 1955. Reprinted in [Hintikka, 1969, pp. 9–41].

[Beth, 1962] E. W. Beth. *Formal Methods.* Reidel, Dordrecht, 1962.

[Bocheński, 1970] I. M. Bocheński. *A History of Formal Logic, translated by I. Thomas.* Chelsea Publishing Co, New York, 1970.

[Bolzano, 1837] B. Bolzano. *Wissenschaftslehre.* 1837. Edited and translated by R.George as *Theory of Science*, UCLA Press, Berkeley and Los Angeles, 1972.

[Boole, 1847] G. Boole. *The Mathematical Analysis of Logic.* Macmillan, Barclay and Macmillan, Cambridge, 1847. Also pp. 45–124 of George Boole, *Studies in Logic and Probability*, Open Court, La Salle, IL, 1952.

[Boole, 1854] G. Boole. *An Investigation of the Laws of Thought.* Walton and Maberley, London, 1854. Republished by Open Court, La Salle, IL, 1952.

[Boolos and Jeffrey, 1989] G. S. Boolos and R. C. Jeffrey. *Computability and Logic.* Cambridge University Press, Cambridge, 1989.

[Boolos, 1979] G. Boolos. *The Unprovability of Consistency: An Essay in Modal Logic.* Cambridge University Press, 1979.

[Boolos, 1993] G. Boolos. *The Logic of Provability.* Cambridge University Press, 1993.

[Carnap, 1935] R. Carnap. Ein Gültigkeitskriterium für die Sätze der klassischen Mathematik. *Monatshefte Math und Phys*, 42:163–190, 1935.

[Carnap, 1956] R. Carnap. *Meaning and Necessity.* University of Chicago Press, second edition, 1956.

[Chang and Keisler, 1973] C. C. Chang and H. J. Keisler. *Model Theory.* North-Holland, Amsterdam, 1973.

[Chastain, 1975] C. Chastain. Reference and context. In K. Gunderson, editor, *Minnesota Studies in the Philosophy of Science, VII, Language, Mind and Knowledge*, pages 194–269. University of Minnesota Press, MI, 1975.

[Cherlin, 1976] G. Cherlin. *Model Theoretic Algebra: Selected Topics*, volume 521 of *Lecture Notes in Maths.* Springer, Berlin, 1976.

[Church, 1936] A. Church. A note on the Entscheidungsproblem. *Journal of Symbolic Logic*, 1:40f, 101f, 1936.

[Church, 1956] A. Church. *Introduction to Mathematical Logic, I.* Princeton University Press, Princeton, NJ, 1956.

[Coffa, 1991] J. A. Coffa. *The Semantic Tradition from Kant to Carnap: To the Vienna Station.* Cambridge University Press, Cambridge, 1991.

[Cohen, 1969] P. J. Cohen. Decision procedures for real and p-adic fields. *Comm Pure Appl Math*, 22:131–151, 1969.

[Cohen, 1971] L. J. Cohen. Some remarks on Grice's views about the logical particles of natural language. In Y. Bar-Hillel, editor, *Pragmatics of Natural Languages*, pages 60–68. Reidel, Dordrecht, 1971.

[Cook, 1971] S. A. Cook. The complexity of theorem-proving procedures. In *Proceedings of the Third Annual ACM Symposium on Theory of Computing*, pages 151–158. ACM Press, NY, 1971.

[Craig, 1957a] W. Craig. Linear reasoning. A new form of the Herbrand–Gentzen theorem. *Journal of Symbolic Logic*, 22:250–268, 1957.

[Craig, 1957b] W. Craig. Three uses of the Herbrand–Gentzen theorem in relating model theory and proof theory. *Journal of Symbolic Logic*, 22:269–285, 1957.

[Dalen, 1980] D. van Dalen. *Logic and Structure.* Springer, Berlin, 1980.

[Dedekind, 1888] R. Dedekind. *Was sind und was sollen die Zahlen?* Brunswick, 1888.

[Dedekind, 1967] R. Dedekind. Letter to Keferstein, 1890. In J. Van Heijenoort, editor, *From Frege to Gödel, A Source Book in Mathematical Logic, 1879–1931*, pages 90–103. Harvard University Press, Cambridge, MA, 1967.

[Došen and Schroeder-Heister, 1993] K. Došen and P. Schroeder-Heister, editors. *Substructural Logics.* Oxford University Press, Oxford, 1993.

[Dowty *et al.*, 1981] D. Dowty, R. Wall, and S. Peters. *Introduction to Montague Semantics.* Reidel, Dordrecht, 1981.

[Dummett, 1958/59] M. A. E. Dummett. Truth. *Proc Aristotelian Soc*, 59:141–162, 1958/59. Reprinted in [Strawson, 1967; pp. 49–68].

[Dummett, 1973] M. A. E. Dummett. *Frege: Philosophy of Language.* Duckworth, London, 1973.

[Dummett, 1975] M. A. E. Dummett. What is a theory of meaning? In Samuel Guttenplan, editor, *Mind and Language.* Clarendon Press, Oxford, 1975.

[Dunn and Belnap, 1968] J. M. Dunn and N. D. Belnap. The substitution interpretation of the quantifiers. *Noûs*, 2:177–185, 1968.

[Ebbinghaus and Flum, 1995] H.-D. Ebbinghaus and J. Flum. *Finite model theory.* Springer, Berlin, 1995.

[Ehrenfeucht, 1960] A. Ehrenfeucht. An application of games to the completeness problem for formalized theories. *Fundamenta Math*, 49:129–141, 1960.

[Enderton, 1972] H. B. Enderton. *A Mathematical Introduction to Logic.* Academic Press, New York, 1972.

[Etchemendy, 1990] J. Etchemendy. *The Concept of Logical Consequence.* Harvard University Press, Cambridge MA, 1990.

[Evans, 1980] G. Evans. Pronouns. *Linguistic Inquiry*, 11:337–362, 1980.

[Feferman, 1968a] S. Feferman. Lectures on proof theory. In *Proc Summer School of Logic, Leeds 1967*, Lecture Notes in Mathematics 70, pages 1–109. Springer, Berlin, 1968.

[Feferman, 1968b] S. Feferman. Persistent and invariant formulas for outer extensions. *Compositio Math*, 20:29–52, 1968.

[Feferman, 1969] S. Feferman. Set-theoretical foundations of category theory. In *Reports of the Midwest Category Seminar III*, Lecture Notes in Mathematics 106, pages 201–247. Springer, Berlin, 1969.

[Feferman, 1974] S. Feferman. Applications of many-sorted interpolation theorems. In L. Henkin *et al.*, editor, *Proceedings of the Tarski Symposium, Proc Symposia in Pure Math. XXV*, pages 205–223. American Mathematical Society, Providence, RI, 1974.

[Feferman, 1984] S. Feferman. Kurt Gödel: conviction and caution. *Philosophia Naturalis*, 21:546–562, 1984.

[Fitch, 1952] F. B. Fitch. *Symbolic Logic.* Ronald Press, New York, 1952.

[Flum, 1975] J. Flum. First-order logic and its extensions. In *ISILC Logic Conference*, Lecture Notes in Mathematics 499, pages 248–307. Springer, Berlin, 1975.

[Fraenkel, 1922] A. Fraenkel. Zu den Grundlagen der Cantor–Zermeloschen Mengenlehre. *Math Annalen*, 86:230–237, 1922.

[Fraïssé, 1954] R. Fraïssé. Sur l'extension aux relations de quelques propriétés des ordres. *Ann Sci École Norm Sup*, 71:363–388, 1954.

[Fraïssé, 1955] R. Fraïssé. Sur quelques classifications des relations, basées sur des isomorphismes restreints. *Alger-Mathématiques*, 2:16–60 and 273–295, 1955.

[Frege, 1879] G. Frege. *Begriffsschrift.* Halle, 1879. Translated in [Heijenoort, 1967, pp. 1–82].

[Frege, 1884] G. Frege. *Die Grundlagen der Arithmetik.* Breslau, 1884. Translated by J. L. Austin, *The Foundations of Arithmetic*, 2nd edn. , Blackwell, Oxford, 1953.

[Frege, 1891] G. Frege. *Funktion und Begriff.* Jena, 1891. Also in [Frege, 1967, pp. 125–142] and translated in [Frege, 1952].

[Frege, 1893] G. Frege. *Grundgesetze der Arithmetik I.* Jena, 1893. Partial translation with introduction by M. Furth, *The Basic Laws of Arithmetic*, University California Press, Berkeley, 1964.

[Frege, 1906] G. Frege. Über die Grundlagen der Geometrie. *Jahresbericht der Deutschen Mathematiker-Vereinigung*, 15:293–309, 377–403 and 423–430, 1906. Translated in [Frege, 1971].

[Frege, 1912] G. Frege. *Anmerkungen zu: Philip E. B. Jourdain.* The development of the theories of mathematical logic and the principles of mathematics, 1912. In [Frege, 1967, pp. 334–341].

[Frege, 1952] G. Frege. Translations from the Philosophical Writings of Gottlob Frege. Blackwell, Oxford, 1952.

[Frege, 1967] G. Frege. Kleine Schriften. Georg Olms Verlagsbuchhandlung, Hildesheim, 1967.

[Frege, 1971] G. Frege. *On the Foundations of Geometry and Formal Theories of Arithmetic.* Yale University Press, New Haven, 1971. Translated with introduction by E. W. Kluge.

[Frege and Hilbert, 1899–1900] G. Frege and D. Hilbert. Correspondence leading to 'On the foundations of geometry', 1899–1900. In [Frege, 1967; pp. 407–418], translated in [Frege, 1971; pp. 6–21].

[Gallier, 1986] J. H. Gallier. *Logic for Computer Science: foundations of Automatic Theorem Proving.* Harper and Row, 1986.

[Gandy, 1974] R. O. Gandy. Set-theoretic functions for elementary syntax. In T. J. Jech, editor, *Axiomatic Set Theory II*, pages 103–126. American Mathematical Society, Providence, RI, 1974.

[Garey and Johnson, 1979] M. R. Garey and D. S. Johnson. *Computers and Intractability.* W. H. Freeman, San Francisco, 1979.

[Gentzen, 1934] G. Gentzen. Untersuchungen über das logische Schliessen. *Math Zeitschrift*, 39:176–210 and 405–431, 1934.

[Girard, 1987] J.-Y. Girard. Linear logic. *Theoretical Computer Science*, 50:1–102, 1987.

[Girard, 1995] J.-Y. Girard. Linear logic: its syntax and semantics. In J.-Y. Girard et al., editor, *Advances in Linear Logic*, pages 1–42. Cambridge University Press, 1995.

[Gödel, 1930] K. Gödel. Die Vollständigkeit der Axiome des logischen Funktionenkalküls. *Monatshefte für Mathematik und Physik*, 37:349–360, 1930. Translated in [Gödel, 1986, pp. 102–123] and [Heijenoort, 1967, pp. 582–591].

[Gödel, 1931a] K. Gödel. Eine Eigenschaft der Realisierungen des Aussagenkalküls. *Ergebnisse Math Kolloq*, 3:20–21, 1931. Translated in [Gödel, 1986, pp. 238–241].

[Gödel, 1931b] K. Gödel. Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I. *Monatshefte für Mathematik und Physik*, 38:173–198, 1931. Translated in [Gödel, 1986, pp. 144–195] and [Heijenoort, 1967, pp. 596–616].

[Gödel, 1947] K. Gödel. What is Cantor's continuum problem? *American Mathematical Monthly*, 54:515–525, 1947. Revised and expanded version in [Gödel, 1990, pp. 254–270].

[Gödel, 1951] K. Gödel. Russell's mathematical logic. In P. A. Schilpp, editor, *The Philosophy of Bertrand Russell*, pages pp. 123–153. Tudor Publ. Co, New York, 1951. Also in [Gödel, 1990, pp. 119–141].

[Gödel, 1986] K. Gödel. *Collected Works. Volume I.* Oxford University Press, New York, 1986. Edited by S. Feferman *et al.*

[Gödel, 1990] K. Gödel. *Collected Works. Volume II.* Oxford University Press, New York, 1990. Edited by S. Feferman *et al.*

[Goldblatt, 1982] R. Goldblatt. *Axiomatizing the Logic of Computer Programming.* Lecture Notes in Computer Science, 130, Springer, Berlin, 1982.

[Goldfarb, 1979] W. D. Goldfarb. Logic in the twenties: the nature of the quantifier. *Journal of Symbolic Logic*, 44:351–368, 1979.

[Goldson, Reeves and Bornat, 1993] D. Goldson, S. Reeves and R. Bornat. A review of several programs for the teaching of logic, *Computer Journal*, 36:373–386, 1993.

[Gómez-Torrente, 1996] M. Gómez-Torrente. Tarski on logical consequence, *Notre Dame Journal of Formal Logic*, 37:125–151, 1996.

[Grice, 1975] H. P. Grice. Logic and conversation. In P. Cole *et al.*, editor, *Syntax and Semantics 3, Speech Acts*, pp. 41–58. Academic Press, New York, 1975. Revised version in P. Grice, *Studies in the Way of Words*, Harvard University Press, Cambridge, MA, 1989, pp. 22–40.

[Groenendijk and Stokhof, 1991] J. Groenendijk and M. Stokhof. Dynamic predicate logic, *Linguistics and Philosophy*, 14:39–100, 1991.

[Gurevich, 1984] Y. Gurevich. Toward logic tailored for computational complexity. In M. M. Richter, *et al.*, editors, *Computation and Proof Theory*, Lecture Notes in Mathematics 1104, pp. 175–216, Springer-Verlag, 1984.

[Hammer, 1995] E. M. Hammer. *Logic and Visual Information.* CSLI and FoLLI, Stanford CA, 1995.

[Harel, 1979] D. Harel. *First-order Dynamic Logic.* Lecture Notes in Computer Science, 68. Springer, Berlin, 1979.

[Harnik, 1985] V. Harnik. Stability theory and set existence axioms. *Journal of Symbolic Logic*, 50:123–137, 1985.

[Harnik, 1987] V. Harnik. Set existence axioms for general (not necessarily countable) stability theory. *Annals of Pure and Applied Logic*, 34:231–243, 1987.

[Hasenjaeger, 1953] G. Hasenjaeger. Eine Bemerkung zu Henkins Beweis für die Vollständigkeit des Prädikatenkalküls der ersten Stufe. *Journal of Symbolic Logic*, 18:42–48, 1953.

[Hausdorff, 1914] F. Hausdorff. *Grundzüge der Mengenlehre.* Veit, Leipzig, 1914.

[Heijenoort, 1967] J. van Heijenoort, editor. *From Frege to Gödel, A Source Book in Mathematical Logic, 1879–1931.* Harvard University Press, Cambridge, MA, 1967.

[Heim, 1988] I. Heim. *The Semantics of Definite and Indefinite Noun Phrases in English*, Garland, New York, 1988.

[Henkin, 1949] L. Henkin. The completeness of the first-order functional calculus. *Journal of Symbolic Logic*, 14:159–166, 1949. Reprinted in [Hintikka, 1969].

[Henkin, 1950] L. Henkin. Completeness in the theory of types. *J. Symbolic Logic*, 15:81–91, 1950. Reprinted in [Hintikka, 1969].

[Henkin, 1961] L. Henkin. Some remarks on infinitely long formulas. In *Infinitistic Methods: Proc. Symp. on Foundations of Mathematics, Warsaw*, pages 167–183. Pergamon, London, 1961.

[Henkin and Mostowski, 1959] L. Henkin and A. Mostowski. Review of Mal'tsev [1941]. *Journal of Symbolic Logic*, 24:55–57, 1959.

[Herbrand, 1930] J. Herbrand. *Recherches sur la théorie de la démonstration.* PhD thesis, University of Paris, 1930. Translated in [Herbrand, 1971, pp. 44–202].

[Herbrand, 1971] J. Herbrand. *Logical Writings.* Harvard University Press, Cambridge, MA, 1971. Edited by W. D. Goldfarb.

[Hilbert, 1899] D. Hilbert. *Grundlagen der Geometrie.* Teubner, Leipzig, 1899.

[Hilbert, 1923] D. Hilbert. Die logischen Grundlagen der Mathematik. *Math Annalen*, 88:151–165, 1923. Also in [Hilbert, 1970, pp. 178–195].

[Hilbert, 1926] D. Hilbert. Über das Unendliche. *Math Annalen*, 95:161–190, 1926. Translated in [Heijenoort, 1967, pp. 367–392]; partial translation in [Benacerraf and Putnam, 1983, pp. 183–201].

[Hilbert, 1928] D. Hilbert. Die Grundlagen der Mathematik. *Abhandlungen aus dem Math. Seminar der Hamburgischen Universität*, 6:65–85, 1928. Translated in [Heijenoort, 1967, pp. 464–479].

[Hilbert, 1970] D. Hilbert. *Gesammelte Abhandlungen III: Analysis, Grundlagen der Mathematik, Physik, Verschiedenes.* Springer, Berlin, 1970.

[Hilbert and Ackermann, 1928] D. Hilbert and W. Ackermann. *Grundzüge der theoretischen Logik.* Springer, Berlin, 1928.

[Hilbert and Bernays, 1934] D. Hilbert and P. Bernays. *Grundlagen der Mathematik I.* Springer, Berlin, 1934.

[Hilbert and Bernays, 1939] D. Hilbert and P. Bernays. *Grundlagen der Mathematik II.* Springer, Berlin, 1939.

[Hintikka, 1953] J. Hintikka. Distributive normal forms in the calculus of predicates. *Acta Philosophica Fennica*, 6, 1953.

[Hintikka, 1955] J. Hintikka. Form and content in quantification theory. *Acta Philosophica Fennica*, 8:7–55, 1955.

[Hintikka, 1969] J. Hintikka, editor. *The Philosophy of Mathematics.* Oxford University Press, 1969.

[Hintikka, 1973] J. Hintikka. *Logic, Language-games and Information.* Oxford University Press, 1973.

[Hintikka, 1996] J. Hintikka. *The Principles of Mathematics Revisited*, Cambridge University Press, Cambridge, 1996.

[Hodges, 1972] W. Hodges. On order-types of models. *Journal of Symbolic Logic*, 37:69f, 1972.

[Hodges, 1977] W. Hodges. *Logic.* Penguin Books, Harmondsworth, Middx, 1977.

[Hodges, 1985/86] W. Hodges. Truth in a structure, *Proceedings of Aristotelian Society*, 86:135–151, 1985/6.

[Hodges, 1993a] W. Hodges. *Model Theory*, Cambridge University Press, Cambridge, 1993.

[Hodges, 1993b] W. Hodges. Logical features of Horn clauses. In *Handbook of Logic in Artificial Intelligence and Logic Programming, Volume 1: Logical Foundations*, D. M. Gabbay, C. J. Hogger and J. A. Robinson, editors. pages 449–503. Clarendon Press, Oxford, 1993.

[Hodges, 1997a] W. Hodges. *A Shorter Model Theory*, Cambridge University Press, Cambridge, 1997.

[Hodges, 1997b] W. Hodges. Compositional semantics for a language of imperfect information, *Logic Journal of the IGPL*, 5:539–563, 1997.

[Huntington, 1904] E. V. Huntington. *The Continuum and Other Types of Serial Order, with an Introduction to Cantor's Transfinite Numbers.* Harvard University Press, Cambridge, MA, 1904.

[Jeffrey, 1967] R. C. Jeffrey. *Formal Logic: its Scope and Limits.* McGraw-Hill, New York, 1967.

[Johnson-Laird and Byrne, 1991] P. N. Johnson-Laird and R. M. J. Byrne. *Deduction.* Lawrence Erlbaum Associates, Hove, 1991.

[Johnstone, 1977] P. T. Johnstone. *Topos Theory.* Academic Press, London, 1977.

[Kalish and Montague, 1964] D. Kalish and R. Montague. *Logic: Techniques of Formal Reasoning.* Harcourt, Brace and World, New York, 1964.

[Kalmár, 1934/5] L. Kalmár. Über die Axiomatisierbarkeit des Aussagenkalküls. *Acta Scient. Math. Szeged*, 7:222–243, 1934/5.

[Kamp, 1971] H. Kamp. Formal properties of 'Now'. *Theoria*, 37:227–273, 1971.

[Kamp, 1981] H. Kamp. A theory of truth and semantic representation. In J. A. G. Groenendijk *et al.*, editor, *Formal Methods in the Study of Language*, pages 277–322. Math Centrum, Amsterdam, 1981.

[Kamp and Reyle, 1993] H. Kamp and U. Reyle. *From Discourse to Logic*, Kluwer, Dordrecht, 1993.

[Kaplan, 1966] D. Kaplan. What is Russell's theory of descriptions? In *Proceedings of Internat Colloquium on Logic, Physical Reality and History, Denver, 1966*, pages 227–244. Plenum, New York, 1966. Reprinted in [Pears, 1972, pp. 227–244].

[Karp, 1965] C. Karp. Finite quantifier equivalence. In J. Addison *et al.*, editor, *The Theory of Models.* North-Holland, Amsterdam, 1965.

[Kempson, 1995] R. Kempson, editor. *Bulletin of the IGPL*, volume 3 numbers 2, 3: Special Issue on Deduction and Language, 1995.

[Kleene, 1943] S.C. Kleene. Recursive predicates and quantifiers. *Trans Amer Math Soc*, 53:41–73, 1943.

[Kleene, 1952] S. C. Kleene. *Introduction to Metamathematics.* North-Holland, Amsterdam, 1952.

[Klenk, 1976] V. Klenk. Intended models and the Löwenheim–Skolem theorem. *J. Philos. Logic*, 5:475–489, 1976.

[Kneale, 1956] W. Kneale. The province of logic. In H. D. Lewis, editor, *Contemporary British Philosophy, 3rd Series*, pages 237–261. George Allen and Unwin, London, 1956.

[Kowalski, 1979] R. Kowalski. *Logic for problem solving*, North-Holland, New York, 1979.

[Kreisel, 1967] G. Kreisel. Informal rigour and completeness proofs. In Lakatos, editor, *Problems in the Philosophy of Mathematics*, pages 138–157. North-Holland, Amsterdam, 1967. Partially reprinted in [Hintikka, 1969, pp. 78–94].

[Kreisel and Krivine, 1967] G. Kreisel and J. L. Krivine. *Elements of Mathematical Logic (Model Theory).* North-Holland, Amsterdam, 1967.

[Kripke, 1976] S. Kripke. Is there a problem about substitutional quantification? In G. Evans and J. McDowell, editors, *Truth and Meaning: Essays in Semantics*, pages 325–419. Clarendon Press, Oxford, 1976.

[Kronecker, 1882] L. Kronecker. Grundzüge einer arithmetischen Theorie der algebraischen Grössen. *Crelle's Journal*, 92:1–122, 1882.

[Lakoff, 1972] G. Lakoff. Linguistics and natural logic. In D. Davidson and G. Harman, editors, *Semantics of Natural Languages*, pages 545–665. Reidel, Dordrecht, 1972.

[Langford, 1927] C. H. Langford. Some theorems on deducibility. *Annals of Math*, 28:16–40, 1927.

[Leisenring, 1969] A. C. Leisenring. *Mathematical Logic and Hilbert's ϵ-symbol.* Gordon and Breach, New York, 1969.

[Lemmon, 1965] E. J. Lemmon. *Beginning Logic.* Nelson, London, 1965.

[Levy, 1965] A. Levy. A hierarchy of formulas in set theory. *Memoirs of the American Mathematical Society*, 57, 1965.

[Levy, 1979] A. Levy. *Basic Set Theory.* Springer, New York, 1979.

[Lindström, 1969] P. Lindström. On extensions of elementary logic. *Theoria*, 35:1–11, 1969.

[Lorenzen, 1961] P. Lorenzen. Ein dialogisches Konstruktivitätskriterium. In *Infinitistic Methods, Proc of a Symp on Foundations of Mathematics, Warsaw*, pages 193–200, Pergamon, London, 1961.

[Lorenzen, 1962] P. Lorenzen. *Metamathematik.* Bibliographisches Institut, Mannheim, 1962.

[Lorenzen and Schwemmer, 1975] P. Lorenzen and O. Schwemmer. *Konstruktive Logic, Ethik und Wissenschaftstheorie.* Bibliographisches Institut, Mannheim, 1975.

[Löwenheim, 1915] L. Löwenheim. Über Möglichkeiten im Relativkallkül. *Math Annalen*, 76:447–470, 1915. Translated in [Heijenoort, 1967, pp . 228–251].

[Łukasiewicz and Tarski, 1930] J. Łukasiewicz and A. Tarski. Untersuchungen über den Aussagenkalkül. *Comptes Rendus des séances de la Société des Sciences et des Lettres de Varsovie*, 23 cl. iii:30–50, 1930. Translated in [Tarski, 1983, pp. 38–59].

[Mal'tsev, 1936] A. I. Mal'tsev. Untersuchungen aus dem Gebiete der Mathematischen Logik. *Mat Sbornik*, 1:323–336, 1936. Translated in [Mal'tsev, 1971, pp. 1–14].

[Mal'tsev, 1941] A. I. Mal'tsev. On a general method for obtaining local theorems in group theory (Russian). *Ivanov Gos. Ped. Inst. Uc. Zap. Fiz.-Mat. Fak.*, 1:3–9, 1941. Translated in [Mal'tsev, 1971, pp. 15–21].

[Mal'tsev, 1971] A. I. Mal'cev. *The Metamathematics of Algebraic Systems; Collected Papers 1936–1967.* North-Holland, Amsterdam, 1971. Translated and edited by B. F. Wells III.

[Manktelow and Over, 1990] K. I. Manktelow and D. E. Over. *Inference and Understanding*, Routledge, London, 1990.

[Mates, 1965] B. Mates. *Elementary Logic.* Oxford University Press, New York, 1965.

[Members of the Johns Hopkins University, Boston, 1883] Members of the Johns Hopkins University, Boston. *Studies in Logic.* Little, Brown and Co, 1883.

[Mendelson, 1987] E. Mendelson. *Introduction to Mathematical Logic*, Third edition. Van Nostrand, Princeton, NJ, 1964.

[Mitchell, 1883] O. H. Mitchell. On a new algebra of logic. In Members of the Johns Hopkins University, Boston, *Studies in Logic*, pages 72–106. Little, Brown and Co, 1883.

[Montague, 1970] R. Montague. English as a formal language. In B. Visentini *et al.*, editor, *Linguaggi nella Società e nella Tecnica.* Milan, 1970. Also in [Montague, 1974, pp. 188–221].

[Montague, 1973] R. Montague. The proper treatment of quantification in ordinary English. In J. Hintikka *et al.*, editor, *Approaches to Natural Language*. Reidel, Dordrecht, 1973. Also in [Montague, 1974, pp. 247–270].

[Montague, 1974] R. H. Thomason, editor. *Formal Philosophy, Selected Papers of Richard Montague*, Yale University Press, New Haven, 1974.

[Montague and Vaught, 1959] R. Montague and R. L. Vaught. Natural models of set theory. *Fundamenta Math*, 47:219–242, 1959.

[Moore, 1980] G. H. Moore. Beyond first-order logic: the historical interplay between mathematical logic and axiomatic set theory. *History and Philosophy of Logic*, 1:95–137, 1980.

[Morrill, 1994] G. V. Morrill. *Type Logical Grammar: Categorial Logic of Signs*. Kluwer, Dordrecht, 1994.

[Nisbett *et al.*, 1987] R. E. Nisbett, G. T. Fong, D. R Lehman and P. W. Cheng. Teaching reasoning. *Science*, 238:625–631, 1987.

[Padawitz, 1988] P. Padawitz. *Computing in Horn Clause Theories*. Springer, Berlin, 1988.

[Partee, 1978] B. Partee. Bound variables and other anaphors. In D. Waltz, editor, *Tinlap-2, Theoretical Issues in Natural Language Processing*, pages 248–280. Association for Computing Machinery, New York, 1978.

[Peano, 1889] G. Peano. *Arithmetices Principia, Nova Methodo Exposita*. Turin, 1889. Translation in [Heijenoort, 1967, pp. 85–97].

[Pears, 1972] D. F. Pears, editor. *Bertrand Russell. A Collection of Critical Essays*. Anchor Books, Doubleday, New York, 1972.

[Peirce, 1883] C. S. Peirce. A theory of probable inference. Note B. The logic of relatives. In Boston Members of the Johns Hopkins University, editor, *Studies in Logic*. Little, Brown and Co, 1883. Reprinted in [Peirce, 1933, Vol III, pp. 195–209].

[Peirce, 1885] C. S. Peirce. On the algebra of logic. *Amer. J. Math.*, 7:180–202, 1885. Reprinted in [Peirce, 1933, Vol. III, pp. 210–238].

[Peirce, 1902] C. S. Peirce. The simplest mathematics. In C. Hartshorne *et al.*, editor, *Collected Papers of Charles Sanders Peirce*, volume IV, pages 189–262. Harvard University Press, Cambridge, MA, 1902.

[Peirce, 1933] C. S. Peirce. In C. Hartshorne *et al.*, editor, *Collected Papers of Charles Sanders Peirce*. Harvard University Press, Cambridge, MA, 1933.

[Perry, 1977] J. Perry. Frege on demonstratives. *Philosophical Review*, 86:474–497, 1977. Reprinted in P. Yourgram, editor, *Demonstratives*, pages 50–70, Oxford University Press, New York, 1990.

[Popper, 1946/47] K. R. Popper. Logic without assumptions. *Proc. Aristot. Soc*, pages 251–292, 1946/47.

[Post, 1921] E. Post. Introduction to a general theory of elementary propositions. *American Journal of Mathematics*, 43:163–185, 1921. Reprinted in [Heijenoort, 1967, pp. 264–283].

[Prawitz, 1965] D. Prawitz. *Natural Deduction: a Proof-theoretical Study*. Almqvist and Wiksell, Stockholm, 1965.

[Prawitz, 1979] D. Prawitz. Proofs and the meaning and the completeness of the logical constants. In J. Hintikka, I. Niiniluoto, and E. Saarinen, editors, *Essays on Mathematical and Philosophical Logic*, pages 25–40. Reidel, Dordrecht, 1979.

[Prior, 1960] A. N. Prior. The runabout inference ticket. *Analysis*, 21:38–39, 1960. Reprinted in [Strawson, 1967, pp. 129–131].

[Prior, 1962] A. N. Prior. *Formal Logic*. Oxford University Press, 1962.

[Putnam, 1980] H. Putnam. Models and reality. *Journal of Symbolic Logic*, 45:464–482, 1980. Reprinted in [Benacerraf, 1983; pp. 421–444].

[Quine, 1940] W. V. Quine. *Mathematical Logic*. Harvard University Press, Cambridge, MA, 1940. Revised edition 1951.

[Quine, 1950] W. V. Quine. *Methods of Logic*. Holt, New York, 1950.

[Quine, 1970] W. V. Quine. *Philosophy of Logic*. Prentice-Hall, Englewood Cliffs, NJ, 1970.

[Quirk and Greenbaum, 1973] R. Quirk and S. Greenbaum. *A University Grammar of English*. Longman, London, 1973.

[Rasiowa and Sikorski, 1950] H. Rasiowa and R. Sikorski. A proof of the completeness theorem of Gödel. *Fundamenta Math*, 37:193–200, 1950.

[Rasiowa and Sikorski, 1963] H. Rasiowa and R. Sikorski. *The Mathematics of Metamathematics.* Monografie Matematyczne, Polska Akad. Nauk, 1963.

[Reeves and Clarke, 1990] S. Reeves and M. Clarke. *Logic for Computer Science.* Addison-Wesley, 1990.

[Rips, 1994] L. J. Rips. *The Psychology of Proof.* MIT Press, Cambridge Mass., 1994.

[Robinson, 1967] A. Robinson. The metaphysics of the calculus. In Lakatos, editor, *Problems in the Philosophy of Mathematics*, pages 28–40. North-Holland, Amsterdam, 1967. Reprinted in [Hintikka, 1969, pp. 153–163], and in Selected papers of Abraham Robinson, Vol. 2, edited by H. J. Keisler *et al.*. pp. 537–555. Yale University Press, New Haven, 1979.

[Russell, 1905] B. Russell. On denoting. *Mind*, 14:479–493, 1905. Reprinted in [Russell, 1956].

[Russell, 1956] B. Russell. In R. C. Marsh, editor, *Logic and Knowledge, Essays 1901–1950.* George Allen and Unwin, London, 1956.

[Sacks, 1972] G. E. Sacks. *Saturated Model Theory.* Benjamin, Reading, MA, 1972.

[Schmidt, 1938] H. A. Schmidt. Über deduktive Theorien mit mehreren Sorten von Grunddingen. *Math Annalen*, 115:485–506, 1938.

[Schröder, 1895] E. Schröder. *Vorlesungen über die Algebra der Logik*, volume 3. Leipzig, 1895.

[Schütte, 1956] K. Schütte. Ein System des verknüpfenden Schliessens. *Arch. Math.Logik Grundlagenforschung*, 2:55–67, 1956.

[Schütte, 1977] K. Schütte. *Proof Theory.* Springer, Berlin, 1977. Translated by J. N. Crossley.

[Scott, 1964] W. R. Scott. *Group Theory.* Prentice-Hall, Englewood Cliffs, NJ, 1964.

[Shoesmith and Smiley, 1978] D. J. Shoesmith and T. J. Smiley. *Multiple-Conclusion Logic.* Cambridge University Press, 1978.

[Skolem, 1919] T. Skolem. Untersuchungen über die Axiome des Klassenkalküls und über Produktations- und Summationsprobleme, welche gewisse von Aussagen betreffen. *Videnskapsselskapets Skrifter, I. Matem.-naturv. klasse, no 3*, 1919. Reprinted in [Skolem, 1970, pp. 67–101].

[Skolem, 1920] T. Skolem. Logisch-kombinatorische Untersuchungen über die Erfüllbarkeit oder Beweisbarkeit mathematischer Sätze nebst einem Theoreme über dichte Mengen. *Videnskapsselskapets Skrifter, I. Matem.-Naturv. Klasse 4*, 1920. Reprinted in [Skolem, 1970, pp. 103–136]; partial translation in [Heijenoort, 1967, pp. 252–263].

[Skolem, 1922] T. Skolem. Einige Bemerkungen zur axiomatischen Begründung der Mengenlehre. *Matematikerkongressen i Helsingfors den 4–7 Juli 1922*, 1922. Reprinted in [Skolem, 1970, pp. 137–152]; translation in [Heijenoort, 1967, pp. 290–301].

[Skolem, 1923] T. Skolem. Begründung der elementaren Arithmetik durch die rekurrierende Denkweise ohne Anwendung scheinbarer Veränderlichen mit unendlichem Ausdehnungsbereich. *Videnskapsselskapets Skrifter I, Matem.-naturv. Klasse 6*, 1923. Translation in [Heijenoort, 1967, pp. 303–333].

[Skolem, 1928] T. Skolem. Über die mathematische Logik. *Norsk. Mat. Tidsk.*, 10:125–142, 1928. Reprinted in [Skolem, 1970, pp. 189–206]; translation in [Heijenoort, 1967, pp. 513–524].

[Skolem, 1929] T. Skolem. Über einige Grundlagenfragen der Mathematik. *Skr. Norsk. Akad. Oslo I Mat.-Natur Kl 4*, pages 1–49, 1929. Reprinted in [Skolem, 1970, pp. 227–273].

[Skolem, 1934] T. Skolem. Über die Nichtcharakterisierbarkeit der Zahlenreihe mittels endlich oder abzählbar unendlich vieler Aussagen mit ausschliesslich Zahlenvariablen. *Fundamenta Math*, 23:150–161, 1934. Reprinted in [Skolem, 1970, pp. 355–366].

[Skolem, 1955] T. Skolem. A critical remark on foundational research. *Kongelige Norsk. Vidensk. Forhand. Trondheim*, 28:100–105, 1955. Reprinted in [Skolem, 1970, pp. 581–586].

[Skolem, 1970] T. Skolem. In *Selected Works in Logic.* J. E. Fenstad, editor, Universitetsforlaget, Oslo, 1970.

[Smullyan, 1968] R. Smullyan. *First-Order Logic.* Springer, Berlin, 1968.

[Sneed, 1971] J. D. Sneed. *The Logical Structure of Mathematical Physics.* Reidel, Dordrecht, 1971.

[Stegmüller, 1976] W. Stegmüller. *The Structure and Dynamics of Theories.* Springer, New York, 1976.

[Steiner, 1975] M. Steiner. *Mathematical Knowledge.* Cornell University Press, Ithaca, 1975.

[Stenning *et al.*, 1995] K. Stenning, R. Cox and J. Oberlander. Contrasting the cognitive effects of graphical and sentential logic teaching: reasoning, representation and individual differences, *Language and Cognitive Processes*, 10:333–354, 1995.

[Stevenson, 1973] L. Stevenson. Frege's two definitions of quantification. *Philos. Quarterly*, 23:207–223, 1973.

[Strawson, 1967] P. F. Strawson, editor. *Philosophical Logic.* Oxford University Press, 1967.

[Suppes, 1957] P. Suppes. *Introduction to Logic.* Van Nostrand, Princeton, NJ, 1957.

[Suppes, 1972] P. Suppes. *Axiomatic Set Theory.* Dover, NY, 1972.

[Tarski, 1935] A. Tarski. Der Wahrheitsbegriff in den formalisierten Sprachen, based on a paper in *Ruch. Filozoficzny xii (1930/1)*, 1935. Translated in [Tarski, 1983, pp. 152–278].

[Tarski, 1936] A. Tarski. O pojciu wynikania logicznego. *Przeglad Filozoficzny*, 39:58–68, 1936. Translated as 'On the concept of logical consequence' in [Tarski, 1983, pp. 409–420].

[Tarski, 1954] A. Tarski. Contributions to the theory of models I, II. *Indag. Math*, 16:572–588, 1954.

[Tarski, 1983] A. Tarski. *Logic, Semantics, Metamathematics, Papers from 1923 to 1938.* Hackett, Indianapolis, 1983. Translated by J. H. Woodger with analytical index by J. Corcoran.

[Tarski and Givant, 1987] A. Tarski and S. Givant. *A Formalization of Set Theory without Variables.* American Mathematical Society, Providence RI, 1987.

[Tarski and Vaught, 1956] A. Tarski and R. L. Vaught. Arithmetical extensions of relational systems. *Compositio Math*, 13:81–102, 1956.

[Tarski *et al.*, 1953] A. Tarski, A. Mostowski, and R. M. Robinson. *Undecidable Theories.* North-Holland, Amsterdam, 1953.

[Tennant, 1978] N. W. Tennant. *Natural Logic.* Edinburgh University Press, 1978.

[Thomason, 1970] R. H. Thomason. *Symbolic Logic, An Introduction.* Macmillan, London, 1970.

[Vaught, 1974] R. L. Vaught. Model theory before 1945. In L. Henkin *et al.*, editor, *Proceedings of the Tarski Symposium*, pages 153–172. AMS, Providence, RI, 1974.

[von Neumann, 1925] J. Von Neumann. Eine Axiomatisierung der Mengenlehre. *J. für die Reine und Angew Math*, 154:219–240, 1925. Translated in [Heijenoort, 1967, pp. 393–413].

[Wang, 1952] H. Wang. Logic of many-sorted theories. *Journal of Symbolic Logic*, 17:105–116, 1952.

[Wang, 1970] H. Wang. A survey of Skolem's work in logic, 1970. In [Skolem, 1970, pp. 17–52].

[Wang, 1974] H. Wang. *From Mathematics to Philosophy.* Routledge and Kegan Paul, NY, 1974.

[Wason, 1966] P. C. Wason. Reasoning. In *New Horizons in Psychology*, B. Foss, ed., pages 135–151. Penguin, Harmondsworth, 1966.

[Whitehead and Russell, 1910] A. N. Whitehead and B. Russell. *Principia Mathematica I.* Cambridge University Press, 1910. Up to to 56*, reprinted 1962.

[Wiredu, 1973] J. E. Wiredu. Deducibility and inferability. *Mind*, 82:31–55, 1973.

[Wittgenstein, 1910] L. Wittgenstein. *Tractatus Logico-Philosophicus.* Annalen der Naturphilosophie, 1910. Reprinted with translation by D. F. Pears and B. F. McGuinness, Routledge and Kegan Paul, London, 1961.

[Zermelo, 1908] E. Zermelo. Untersuchungen über die Grundlagen der Mengenlehre I. *Math Annalen*, 65:261–281, 1908. Translated in [Heijenoort, 1967, pp. 199–215].

[Zucker, 1974] J. Zucker. The correspondence between cut-elimination and normalisation. *Annals of Math Logic*, 7:1–156, 1974.