Introduction to Mathematical Logic

FOURTH EDITION

Elliott Mendelson
Queens College of the City University of New York

CHAPMAN & HALL
London · Weinheim · New York · Tokyo · Melbourne · Madras

Published by Chapman & Hall, 2–6 Boundary Row, London SE1 8HN, UK

Chapman & Hall, 2-6 Boundary Row, London SE1 8HN, UK

Chapman & Hall GmbH, Pappelallee 3, 69469 Weinheim, Germany

Chapman & Hall USA, 115 Fifth Avenue, New York, NY 10003, USA

Chapman & Hall Japan, ITP-Japan, Kyowa Building, 3F, 2-2-1 Hirakawacho, Chiyoda-ku, Tokyo 102, Japan

Chapman & Hall Australia, 102 Dodds Street, South Melbourne, Victoria 3205, Australia

Chapman & Hall India, R. Seshadri, 32 Second Main Road, CIT East, Madras 600 035, India

First edition 1964

Second edition 1979

Third edition 1987

Fourth edition 1997

Typeset in 10/12 Times by Scientific Publishing Services (P) Ltd., Madras, India

Printed in Great Britain by Hartnolls Ltd, Bodmin, Cornwall.

ISBN 0 412 80830 7

A catalogue record for this book is available from the British Library

♾ Printed on permanent acid-free text paper, manufactured in accordance with ANSI/NISO Z39.48-1992 and ANSI/NISO Z39.48-1984 (Permanence of Paper).

To Arlene

Contents

Preface

This is a compact introduction to some of the principal topics of mathematical logic. In the belief that beginners should be exposed to the easiest and most natural proofs, I have used free-swinging set-theoretic methods. The significance of a demand for constructive proofs can be evaluated only after a certain amount of experience with mathematical logic has been obtained. If we are to be expelled from 'Cantor's paradise' (as non-constructive set theory was called by Hilbert), at least we should know what we are missing.

The major changes in this new edition are the following.

1. In Chapter 2, a section has been added on logic with empty domains, that is, on what happens when we allow interpretations with an empty domain.
2. In Chapter 4, Section 4.6 has been extended to include an outline of an axiomatic set theory with urelements.
3. The subjects of register machines and random access machines have been dropped from Section 5.5 Chapter 5.
4. An appendix on second-order logic will give the reader an idea of the advantages and limitations of the systems of first-order logic used in Chapters 2–4, and will provide an introduction to an area of much current interest.
5. The exposition has been further streamlined, more exercises have been added, and the bibliography has been revised and brought up to date.

The material of the book can be covered in two semesters, but, for a one-semester course, Chapters 1–3 are quite adequate (omitting, if hurried, Sections 1.5, 1.6 and 2.10–2.16). I have adopted the convention of prefixing a D to any section or exercise that will probably be difficult for a beginner, and an A to any section or exercise that presupposes familiarity with a topic that has not been carefully explained in the text. Bibliographic references are given to the best source of information, which is not always the earliest paper; hence these references give no indication as to priority.

I believe that the essential parts of the book can be read with ease by anyone with some experience in abstract mathematical thinking. There is, however, no specific prerequisite.

This book owes an obvious debt to the standard works of Hilbert and

Bernays (1934; 1939), Kleene (1952), Rosser (1953) and Church (1956). I am grateful to many people for their help and would especially like to thank the following people for their valuable suggestions and criticism: Richard Butrick, James Buxton, Frank Cannonito, John Corcoran, Newton C.A. da Costa, Robert Cowen, Anil Gupta, Eric Hammer, Bill Hart, Stephen Hechler, Arnold Koslow, Byeong-deok Lee, Alex Orenstein, Dev K. Roy, Atsumi Shimojima and Frank Vlach.

Elliott Mendelson
August 1996

Introduction

One of the popular definitions of logic is that it is the analysis of methods of reasoning. In studying these methods, logic is interested in the form rather than the content of the argument. For example, consider the two arguments:

1. All men are mortal. Socrates is a man. Hence, Socrates is mortal.
2. All cats like fish. Silvy is a cat. Hence, Silvy likes fish.

Both have the same form: All A are B. S is an A. Hence, S is a B. The truth or falsity of the particular premisses and conclusions is of no concern to logicians. They want to know only whether the premisses imply the conclusion. The systematic formalization and cataloguing of valid methods of reasoning are a main task of logicians. If the work uses mathematical techniques or if it is primarily devoted to the study of mathematical reasoning, then it may be called *mathematical logic*. We can narrow the domain of mathematical logic if we define its principal aim to be a precise and adequate understanding of the notion of *mathematical proof*.

Impeccable definitions have little value at the beginning of the study of a subject. The best way to find out what mathematical logic is about is to start doing it, and students are advised to begin reading the book even though (or especially if) they have qualms about the meaning and purpose of the subject.

Although logic is basic to all other studies, its fundamental and apparently self-evident character discouraged any deep logical investigations until the late 19th century. Then, under the impetus of the discovery of non-Euclidean geometry and the desire to provide a rigorous foundation for calculus and higher analysis, interest in logic revived. This new interest, however, was still rather unenthusiastic until, around the turn of the century, the mathematical world was shocked by the discovery of the paradoxes – that is, arguments that lead to contradictions. The most important paradoxes are described here.

1. *Russell's paradox* (1902). By a set, we mean any collection of objects – for example, the set of all even integers or the set of all saxophone players in Brooklyn. The objects that make up a set are called its members or

elements. Sets may themselves be members of sets; for example, the set of all sets of integers has sets as its members. Most sets are not members of themselves; the set of cats, for example, is not a member of itself because the set of cats is not a cat. However, there may be sets that do belong to themselves – for example, the set of all sets. Now, consider the set A of all those sets X such that X is not a member of X. Clearly, by definition, A is a member of A if and only if A is not a member of A. So, if A is a member of A, then A is also not a member of A; and if A is not a member of A, then A is a member of A. In any case, A is a member of A and A is not a member of A.

2. *Cantor's paradox* (1899). This paradox involves the theory of cardinal numbers and may be skipped by those readers having no previous acquaintance with that theory. The cardinal number $\overline{\overline{Y}}$ of a set Y is a measure of the size of the set; $\overline{\overline{Y}} = \overline{\overline{Z}}$ if and only if Y is equinumerous with Z (that is, there is a one–one correspondence between Y and Z). We define $\overline{\overline{Y}} \leqslant \overline{\overline{Z}}$ to mean that Y is equinumerous with a subset of Z; by $\overline{\overline{Y}} < \overline{\overline{Z}}$ we mean $\overline{\overline{Y}} \leqslant \overline{\overline{Z}}$ and $\overline{\overline{Y}} \neq \overline{\overline{Z}}$. Cantor proved that, if $\mathscr{P}(Y)$ is the set of all subsets of Y, then $\overline{\overline{Y}} < \overline{\overline{\mathscr{P}(Y)}}$. Let V be the universal set – that is, the set of all sets. Now, $\mathscr{P}(V)$ is a subset of V; so it follows easily that $\overline{\overline{\mathscr{P}(V)}} \leqslant \overline{\overline{V}}$. On the other hand, by Cantor's theorem, $\overline{\overline{V}} < \overline{\overline{\mathscr{P}(V)}}$. Bernstein's theorem asserts that, if $\overline{\overline{Y}} \leqslant \overline{\overline{Z}}$ and $\overline{\overline{Z}} \leqslant \overline{\overline{Y}}$, then $\overline{\overline{Y}} = \overline{\overline{Z}}$. Hence, $\overline{\overline{V}} = \overline{\overline{\mathscr{P}(V)}}$, contradicting $\overline{\overline{V}} < \overline{\overline{\mathscr{P}(V)}}$.
3. *Burali-Forti's paradox* (1897). This paradox is the analogue in the theory of ordinal numbers of Cantor's paradox and requires familiarity with ordinal number theory. Given any ordinal number, there is a still larger ordinal number. But the ordinal number determined by the set of all ordinal numbers is the largest ordinal number.
4. *The liar paradox*. A man says, 'I am lying', If he is lying, then what he says is true and so he is not lying. If he is not lying, then what he says is true, and so he is lying. In any case, he is lying and he is not lying.†
5. *Richard's paradox* (1905). Some phrases of the English language denote real numbers; for example, 'the ratio between the circumference and diameter of a circle' denotes the number π. All the phrases of the English language can be enumerated in a standard way: order all phrases that have k letters lexicographically (as in a dictionary) and then place all phrases with k letters before all phrases with a larger number of letters. Hence, all phrases of the English language that denote real numbers can

†The Cretan 'paradox', known in antiquity, is similar to the liar paradox. The Cretan philosopher Epimenides said, 'All Cretans are liars'. If what he said is true, then, since Epimenides is a Cretan, it must be false. Hence, what he said is false. Thus, there must be some Cretan who is not a liar. This is not logically impossible; so we do not have a genuine paradox. However, the fact that the utterance by Epimenides of that false sentence could imply the existence of some Cretan who is not a liar is rather unsettling.

be enumerated merely by omitting all other phrases in the given standard enumeration. Call the nth real number in this enumeration the nth Richard number. Consider the phrase: 'the real number whose nth decimal place is 1 if the nth decimal place of the nth Richard number is not 1, and whose nth decimal place is 2 if the nth decimal place of the nth Richard number is 1.' This phrase defines a Richard number – say, the kth Richard number; but, by its definition, it differs from the kth Richard number in the kth decimal place.

6. *Berry's paradox* (1906). There are only a finite number of symbols (letters, punctuation signs, etc.) in the English language. Hence, there are only a finite number of English expressions that contain fewer than 200 occurrences of symbols (allowing repetitions). There are, therefore, only a finite number of positive integers that are denoted by an English expression containing fewer than 200 occurrences of symbols. Let k be *the least positive integer that is not denoted by an English expression containing fewer than 200 occurrences of symbols*. The italicized English phrase contains fewer than 200 occurrences of symbols and denotes the integer k.
7. *Grelling's paradox* (1908). An adjective is called *autological* if the property denoted by the adjective holds for the adjective itself. An adjective is called *heterological* if the property denoted by the adjective does not apply to the adjective itself. For example, 'polysyllabic' and 'English' are autological, whereas 'monosyllabic' and 'French' are heterological. Consider the adjective 'heterological'. If 'heterological' is heterological, then it is not heterological. If 'heterological' is not heterological, then it is heterological. In either case, 'heterological' is both heterological and not heterological.
8. *Löb's paradox* (1955). Let A be any sentence. Let B be the sentence: 'If this sentence is true, then A'. So, B asserts: 'If B is true, then A'. Now consider the following argument: Assume B is true; then, by B, since B is true, A holds. This argument shows that, if B is true, then A. But this is exactly what B asserts. Hence, B is true. Therefore, by B, since B is true, A is true. Thus, every sentence is true.

All of these paradoxes are genuine in the sense that they contain no obvious logical flaws. The logical paradoxes (1–3) involve only notions from the theory of sets, whereas the semantic paradoxes (4–8) also make use of concepts like 'denote', 'true' and 'adjective', which need not occur within our standard mathematical language. For this reason, the logical paradoxes are a much greater threat to a mathematician's peace of mind than the semantic paradoxes.

Analysis of the paradoxes has led to various proposals for avoiding them. All of these proposals are restrictive in one way or another of the 'naive' concepts that enter into the derivation of the paradoxes. Russell noted the self-reference present in all the paradoxes and suggested that every object

must have a definite non-negative integer as its 'type'. Then an expression 'x is a member of the set y' is to be considered *meaningful* if and only if the type of y is one greater than the type of x.

This approach, known as the theory of types and systematized and developed in *Principia Mathematica* Whitehead and Russell (1910–13), is successful in eliminating the known paradoxes,† but it is clumsy in practice and has certain other drawbacks as well. A different criticism of the logical paradoxes is aimed at their assumption that, for every property $P(x)$, there exists a corresponding set of all objects x that satisfy $P(x)$. If we reject this assumption, then the logical paradoxes are no longer derivable.‡ It is necessary, however, to provide new postulates that will enable us to prove the existence of those sets that are needed by the practising mathematician. The first such axiomatic set theory was invented by Zermelo (1908). In Chapter 4 we shall present an axiomatic theory of sets that is a descendant of Zermelo's system (with some new twists given to it by von Neumann, R. Robinson, Bernays, and Gödel). There are also various hybrid theories combining some aspects of type theory and axiomatic set theory – for example, Quine's system NF.

A more radical interpretation of the paradoxes has been advocated by Brouwer and his intuitionist school (see Heyting, 1956). They refuse to accept the universality of certain basic logical laws, such as the law of excluded middle: P or not-P. Such a law, they claim, is true for finite sets, but it is invalid to extend it on a wholesale basis to all sets. Likewise, they say it is invalid to conclude that 'There exists an object x such that not-$P(x)$' follows from the negation of 'For all x, $P(x)$'; we are justified in asserting the existence of an object having a certain property only if we know an effective method for constructing (or finding) such an object. The paradoxes are not derivable (or even meaningful) if we obey the intuitionist strictures, but so are many important theorems of everyday mathematics, and, for this reason, intuitionism has found few converts among mathematicians.

Whatever approach one takes to the paradoxes, it is necessary first to examine the language of logic and mathematics to see what symbols may be used, to determine the ways in which these symbols are put together to form terms, formulas, sentences and proofs, and to find out what can and cannot be proved if certain axioms and rules of inference are assumed. This is one of the tasks of mathematical logic, and, until it is done, there is no basis for

†Russells's paradox, for example, depends on the existence of the set A of all sets that are not members of themselves. Because, according to the theory of types, it is meaningless to say that a set belongs to itself, there is no such set A.

‡Russell's paradox then proves that there is no set A of all sets that do not belong to themselves. The paradoxes of Cantor and Burali-Forti show that there is no universal set and no set that contains all ordinal numbers. The semantic paradoxes cannot even be formulated, since they involve notions not expressible within the system.

comparing rival foundations of logic and mathematics. The deep and devastating results of Gödel, Tarski, Church, Rosser, Kleene, and many others have been ample reward for the labour invested and have earned for mathematical logic its status as an independent branch of mathematics.

For the absolute novice a summary will be given here of some of the basic notation, ideas, and results used in the text. The reader is urged to skip these explanations now and, if necessary, to refer to them later on.

A set is a collection of objects.† The objects in the collection are called *elements* or *members* of the set. We shall write '$x \in y$' for the statement that x is a member of y. (Synonymous expressions are 'x belongs to y' and 'y contains x'.) The negation of '$x \in y$' will be written '$x \notin y$'.

By '$x \subseteq y$' we mean that every member of x is also a member of y (synonymously, that x is a *subset* of y, or that x is *included* in y). We shall write '$t = s$' to mean that t and s denote the same object. As usual, '$t \neq s$' is the negation of '$t = s$'. For sets x and y, we assume that $x = y$ if and only if $x \subseteq y$ and $y \subseteq x$ – that is, if and only if x and y have the same members. A set x is called a *proper* subset of a set y, written '$x \subset y$' if $x \subseteq y$ but $x \neq y$. (The notation $x \not\subseteq y$ is often used instead of $x \subset y$.)

The *union* $x \cup y$ of sets x and y is defined to be the set of all objects that are members of x or y or both. Hence, $x \cup x = x$, $x \cup y = y \cup x$, and $(x \cup y) \cup z = x \cup (y \cup z)$. The *intersection* $x \cap y$ is the set of objects that x and y have in common. Therefore, $x \cap x = x$, $x \cap y = y \cap x$, and $(x \cap y) \cap z = x \cap (y \cap z)$. Moreover, $x \cap (y \cup z) = (x \cap y) \cup (x \cap z)$ and $x \cup (y \cap z) = (x \cup y) \cap (x \cup z)$. The *relative complement* $x - y$ is the set of members of x that are not members of y. We also postulate the existence of the *empty set* (or *null set*) $\emptyset$ – that is, a set that has no members at all. Then $x \cap \emptyset = \emptyset$, $x \cup \emptyset = x$, $x - \emptyset = x$, $\emptyset - x = \emptyset$, and $x - x = \emptyset$. Sets x and y are called *disjoint* if $x \cap y = \emptyset$.

Given any objects $b_1, \ldots, b_k$, the set that contains $b_1, \ldots, b_k$ as its only members is denoted $\{b_1, \ldots, b_k\}$. In particular, $\{x, y\}$ is a set having x and y as its only members and, if $x \neq y$, is called the *unordered pair* of x and y. The set $\{x, x\}$ is identical with $\{x\}$ and is called the *unit set* of x. Notice that $\{x, y\} = \{y, x\}$. By $\langle b_1, \ldots, b_k \rangle$ we mean the *ordered k-tuple* of $b_1, \ldots, b_k$. The basic property of ordered k-tuples is that $\langle b_1, \ldots, b_k \rangle = \langle c_1, \ldots, c_k \rangle$ if and only if $b_1 = c_1$, $b_2 = c_2, \ldots, b_k = c_k$. Thus, $\langle b_1, b_2 \rangle = \langle b_2, b_1 \rangle$ if and only if $b_1 = b_2$. Ordered 2-tuples are called *ordered pairs*. The ordered 1-tuple $\langle b \rangle$ is taken to be b itself. If X is a set and k is a positive integer, we denote by X^k the set of all ordered k-tuples $\langle b_1, \ldots, b_k \rangle$ of elements $b_1, \ldots, b_k$ of X. In

†Which collections of objects form sets will not be specified here. Care will be exercised to avoid using any ideas or procedures that may lead to the paradoxes; all the results can be formalized in the axiomatic set theory of Chapter 4. The term 'class' is sometimes used as a synonym for 'set', but it will be avoided here because it has a different meaning in Chapter 4. If a property $P(x)$ does determine a set, that set is often denoted $\{x \mid P(x)\}$.

particular, X^1 is X itself. If Y and Z are sets, then by $Y \times Z$ we denote the set of all ordered pairs $\langle y, z \rangle$ such that $y \in Y$ and $z \in Z$. $Y \times Z$ is called the *Cartesian product* of Y and Z.

An *n-place relation* (or a *relation with n arguments*) on a set X is a subset of X^n – that is, a set of ordered n-tuples of elements of X. For example, the 3-place relation of betweenness for points on a line is the set of all 3-tuples $\langle x, y, z \rangle$ such that the point x lies between the points y and z. A 2-place relation is called a *binary* relation; for example, the binary relation of fatherhood on the set of human beings is the set of all ordered pairs $\langle x, y \rangle$ such that x and y are human beings and x is the father of y. A 1-place relation on X is a subset of X and is called a *property* on X.

Given a binary relation R on a set X, the *domain* of R is defined to be the set of all y such that $\langle y, z \rangle \in R$ for some z; the *range* of R is the set of all z such that $\langle y, z \rangle \in R$ for some y; and the *field* of R is the union of the domain and range of R. The *inverse* relation R^{-1} of R is the set of all ordered pairs $\langle y, z \rangle$ such that $\langle z, y \rangle \in R$. For example, the domain of the relation $<$ on the set ω of non-negative integers† is ω, its range is $\omega - \{0\}$, and the inverse of $<$ is $>$. Notation: Very often xRy is written instead of $\langle x, y \rangle \in R$. Thus, in the example just given, we usually write $x < y$ instead of $\langle x, y \rangle \in <$.

A binary relation R is said to be *reflexive* if xRx for all x in the field of R; R is *symmetric* if xRy implies yRx; and R is *transitive* if xRy and yRz imply xRz. Examples: The relation $\leqslant$ on the set of integers is reflexive and transitive but not symmetric. The relation 'having at least one parent in common' on the set of human beings is reflexive and symmetric, but not transitive.

A binary relation that is reflexive, symmetric and transitive is called an *equivalence relation*. Examples of equivalence relations are: (1) the *identity relation* I_X on a set X, consisting of all pairs $\langle x, x \rangle$, where $x \in X$; (2) given a fixed positive integer n, the relation $x \equiv y \pmod{n}$, which holds when x and y are integers and $x - y$ is divisible by n; (3) the congruence relation on the set of triangles in a plane; (4) the similarity relation on the set of triangles in a plane. Given an equivalence relation R whose field is X, and given any $y \in X$, define $[y]$ as the set of all z in X such that yRz. Then $[y]$ is called the *R- equivalence class* of y. Clearly, $[u] = [v]$ if and only if uRv. Moreover, if $[u] \neq [v]$, then $[u] \cap [v] = \emptyset$; that is, different R-equivalence classes have no elements in common. Hence, the set X is completely partitioned into the R-equivalence classes. In example (1) above, the equivalence classes are just the unit sets $\{x\}$, where $x \in X$. In example (2), there are n equivalence classes, the kth equivalence class $(k = 0, 1, \ldots, n-1)$ being the set of all integers that leave the remainder k upon division by n.

A *function* f is a binary relation such that $\langle x, y \rangle \in f$ and $\langle x, z \rangle \in f$ imply $y = z$. Thus, for any element x of the domain of a function f, there is a unique y such that $\langle x, y \rangle \in f$; this unique y is denoted $f(x)$. If x is in the

†ω will also be referred to as the set of *natural numbers*.

domain of f, then $f(x)$ is said to be *defined*. A function f with domain X and range Y is said to be a function from X *onto* Y. If f is a function from X onto a subset of Z, then f is said to be a function from X *into* Z. For example, if the domain of f is the set of integers and $f(x) = 2x$ for every integer x, then f is a function from the set of integers onto the set of even integers, and f is a function from the set of integers into the set of integers. A function whose domain consists of n-tuples is said to be a *function of n arguments*. A *total function of n arguments on a set X* is a function f whose domain is X^n. It is customary to write $f(x_1, \ldots, x_n)$ instead of $f(\langle x_1, \ldots, x_n \rangle)$, and we refer to $f(x_1, \ldots, x_n)$ as the *value* of f for the *arguments* $x_1, \ldots, x_n$. A *partial* function of n arguments on a set X is a function whose domain is a subset of X^n. For example, ordinary division is a partial, but not total, function of two arguments on the set of integers, since division by 0 is not defined. If f is a function with domain X and range Y, then the *restriction* f_Z of f to a set Z is the function $f \cap (Z \times Y)$. Then $f_Z(u) = v$ if and only if $u \in Z$ and $f(u) = v$. The *image* of the set Z under the function f is the range of f_Z. The *inverse image* of a set W under the function f is the set of all u in the domain of f such that $f(u) \in W$. We say that f *maps* X onto (into) Y if X is a subset of the domain of f and the image of X under f is (a subset of) Y. By an *n-place operation* (or *operation with n arguments*) on a set X we mean a function from X^n into X. For example, ordinary addition is a binary (i.e., 2-place) operation on the set of natural numbers $\{0, 1, 2, \cdots\}$. But ordinary subtraction is not a binary operation on the set of natural numbers.

The *composition* $f \circ g$ (sometimes denoted fg) of functions f and g is the function such that $(f \circ g)(x) = f(g(x))$; $(f \circ g)(x)$ is defined if and only if $g(x)$ is defined and $f(g(x))$ is defined. For example, if $g(x) = x^2$ and $f(x) = x + 1$ for every integer x, then $(f \circ g)(x) = x^2 + 1$ and $(g \circ f)(x) = (x + 1)^2$. Also, if $h(x) = -x$ for every real number x and $f(x) = \sqrt{x}$ for every non-negative real number x, then $(f \circ h)(x)$ is defined only for $x \leqslant 0$, and, for such x, $(f \circ h)(x) = \sqrt{-x}$. A function f such that $f(x) = f(y)$ implies $x = y$ is called a *1–1* (*one–one*) function. For example, the identity relation I_X on a set X is a 1–1 function, since $I_X(y) = y$ for every $y \in X$; the function g with domain ω, such that $g(x) = 2x$ for every $x \in \omega$, is 1–1; but the function h whose domain is the set of integers and such that $h(x) = x^2$ for every integer x is not 1–1, since $h(-1) = h(1)$. Notice that a function f is 1–1 if and only if its inverse relation f^{-1} is a function. If the domain and range of a 1–1 function f are X and Y, then f is said to be a 1 – 1 (*one–one*) *correspondence between X and Y*; then f^{-1} is a 1–1 correspondence between Y and X, and $(f^{-1} \circ f) = I_X$ and $(f \circ f^{-1}) = I_Y$. If f is a 1–1 correspondence between X and Y and g is a 1–1 correspondence between Y and Z, then $g \circ f$ is a 1–1 correspondence between X and Z. Sets X and Y are said to be *equinumerous* (written $X \cong Y$) if and only if there is a 1–1 correspondence between X and Y. Clearly, $X \cong X$, $X \cong Y$ implies $Y \cong X$, and $X \cong Y$ and $Y \cong Z$ implies $X \cong Z$. It is somewhat harder to show that, if $X \cong Y_1 \subseteq Y$ and $Y \cong X_1 \subseteq X$,

then $X \cong Y$ (see Bernstein's theorem in Chapter 4). If $X \cong Y$, one says that X and Y *have the same cardinal number*, and if X is equinumerous with a subset of Y but Y is not equinumerous with a subset of X, one says that the cardinal number of X *is smaller than* the cardinal number of Y.[†]

A set X is *denumerable* if it is equinumerous with the set of positive integers. A denumerable set is said to have cardinal number $\aleph_0$, and any set equinumerous with the set of all subsets of a denumerable set is said to have the cardinal number $2^{\aleph_0}$ (or to have the *power of the continuum*). A set X is *finite* if it is empty or if it is equinumerous with the set $\{1, 2, \ldots, n\}$ of all positive integers that are less than or equal to some positive integer n. A set that is not finite is said to be *infinite*. A set is *countable* if it is either finite or denumerable. Clearly, any subset of a denumerable set is countable. A *denumerable sequence* is a function s whose domain is the set of positive integers; one usually writes s_n instead of $s(n)$. A *finite sequence* is a function whose domain is the empty set or $\{1, 2, \ldots, n\}$ for some positive integer n.

Let $P(x, y_1, \ldots, y_k)$ be some relation on the set of non-negative integers. In particular, P may involve only the variable x and thus be a property. If $P(0, y_1, \ldots, y_k)$ holds, and, if, for every n, $P(n, y_1, \ldots, y_k)$ implies $P(n+1, y_1, \ldots, y_k)$, then $P(x, y_1, \ldots, y_k)$ is true for all non-negative integers x (*principle of mathematical induction*). In applying this principle, one usually proves that, for every n, $P(n, y_1, \ldots, y_K)$ implies $P(n+1, y_1, \ldots, y_k)$ by assuming $P(n, y_1, \ldots, y_k)$ and then deducing $P(n+1, y_1, \ldots, y_k)$; in the course of this deduction, $P(n, y_1, \ldots, y_k)$ is called the *inductive hypothesis*. If the relation P actually involves variables $y_1, \ldots, y_k$ other than x, then the proof is said to proceed by *induction on* x. A similar induction principle holds for the set of integers greater than some fixed integer j. An example is: to prove by mathematical induction that the sum of the first n odd integers $1 + 3 + 5 + \ldots + (2n-1)$ is n^2, first show that $1 = 1^2$ (that is, $P(1)$), and then, that if $1 + 3 + 5 + \ldots + (2n-1) = n^2$, then $1 + 3 + 5 + \ldots + (2n-1) + (2n+1) = (n+1)^2$ (that is, if $P(n)$ then $P(n+1)$). From the principle of mathematical induction one can prove the *principle of complete induction*: If, for every non-negative integer x the assumption that $P(u, y_1, \ldots, y_k)$ is true for all $u < x$ implies that $P(x, y_1, \ldots, y_k)$ holds, then, for all non-negative integers x, $P(x, y_1, \ldots, y_k)$ is true, (Exercise: Show by complete induction that every integer greater than 1 is divisible by a prime number.)

A *partial order* is a binary relation R such that R is transitive and, for every x in the field of R, xRx is false. If R is a partial order, then the relation R' that is the union of R and the set of all ordered pairs $\langle x, x \rangle$, where x is in the field of R, we shall call a *reflexive partial order*; in the literature, 'partial order' is used for either partial order or reflexive partial order. Notice that

[†]One can attempt to define the cardinal number of a set X as the collection $[X]$ of all sets equinumerous with X. However, in certain axiomatic set theories, $[X]$ does not exist, whereas in others $[X]$ exists but is not a set.

(xRy and yRx) is impossible if R is a partial order, whereas (xRy and yRx) implies $x = y$ if R is a reflexive partial order. A (reflexive) *total order* is a (reflexive) partial order such that, for any x and y in the field of R, either $x = y$ or xRy or yRx. Examples: (1) the relation $<$ on the set of integers is a total order, whereas $\leqslant$ is a reflexive total order; (2) the relation $\subset$ on the set of all subsets of the set of positive integers is a partial order but not a total order, whereas the relation $\subseteq$ is a reflexive partial order but not a reflexive total order. If B is a subset of the field of a binary relation R, then an element y of B is called an *R-least element* of B if yRz for every element z of B different from y. A *well-order* (or a *well-ordering relation*) is a total order R such that every non-empty subset of the field of R has an R-least element. Examples: (1) the relation $<$ on the set of non-negative integers is a well-order; (2) the relation $<$ on the set of non-negative rational numbers is a total order but not a well-order; (3) the relation $<$ on the set of integers is a total order but not a well-order. Associated with every well-order R having field X there is a *complete induction principle*: if P is a property such that, for any u in X, whenever all z in X such that zRu have the property P, then u has the property P, then it follows that all members of X have the property P. If the set X is infinite, a proof using this principle is called a proof by *transfinite induction*. One says that a set X *can be well-ordered* if there exists a well-order whose field is X. An assumption that is useful in modern mathematics but about the validity of which there has been considerable controversy is the *well-ordering principle*: every set can be well-ordered. The well-ordering principle is equivalent (given the usual axioms of set theory) to the *axiom of choice*: for any set X of non-empty pairwise disjoint sets, there is a set Y (called a *choice set*) that contains exactly one element in common with each set in X.

Let B be a non-empty set, f a function from B into B, and g a function from B^2 into B. Write x' for $f(x)$ and $x \cap y$ for $g(x, y)$. Then $\langle B, f, g\rangle$ is called a *Boolean algebra* if B contains at least two elements and the following conditions are satisfied:

1. $x \cap y = y \cap x$ for all x and y in B
2. $(x \cap y) \cap z = x \cap (y \cap z)$ for all x, y, z in B
3. $x \cap y' = z \cap z'$ if and only if $x \cap y = x$ for all x, y, z in B.

Let $x \cup y$ stand for $(x' \cap y')'$, and write $x \leqslant y$ for $x \cap y = x$. It is easily proved that $z \cap z' = w \cap w'$ for any w and z in B; we denote the value of $z \cap z'$ by 0. Let 1 stand for $0'$. Then $z \cup z' = 1$ for all z in B. Note also that $\leqslant$ is a reflexive partial order on B, and $\langle B, f, \cup\rangle$ is a Boolean algebra. (The symbols $\cap, \cup, 0, 1$ should not be confused with the corresponding symbols used in set theory and arithmetic.) An *ideal* J in $\langle B, f, g\rangle$ is a non-empty subset of B such that (1) if $x \in J$ and $y \in J$, then $x \cup y \in J$, and (2) if $x \in J$ and $y \in B$, then $x \cap y \in J$. Clearly, $\{0\}$ and B are ideals. An ideal different from B is called a *proper* ideal. A *maximal* ideal is a proper ideal that is included in no other

proper ideal. It can be shown that a proper ideal J is maximal if and only if, for any u in $B, u \in J$ or $u' \in J$. From the axiom of choice it can be proved that every Boolean algebra contains a maximal ideal, or, equivalently, that every proper ideal is included in some maximal ideal. For example, let B be the set of all subsets of a set X; for $Y \in B$, let $Y' = X - Y$, and for Y and Z in B, let $Y \cap Z$ be the ordinary set-theoretic intersection of Y and Z. Then $\langle B, ', \cap \rangle$ is a Boolean algebra. The 0 of B is the empty set $\emptyset$, and 1 is X. For each element u in X, the set J_u of all subsets of X that do not contain u is a maximal ideal. For a detailed study of Boolean algebras, see Sikorski (1960), Halmos (1963) and Mendelson (1970).

The Propositional Calculus

1

1.1 PROPOSITIONAL CONNECTIVES. TRUTH TABLES

Sentences may be combined in various ways to form more complicated sentences. We shall consider only *truth-functional* combinations, in which the truth or falsity of the new sentence is determined by the truth or falsity of its component sentences.

Negation is one of the simplest operations on sentences. Although a sentence in a natural language may be negated in many ways, we shall adopt a uniform procedure: placing a sign for negation, the symbol $\neg$, in front of the entire sentence. Thus, if A is a sentence, then $\neg A$ denotes the negation of A.

The truth-functional character of negation is made apparent in the following *truth table*:

A	$\neg A$
T	F
F	T

When A is true, $\neg A$ is false; when A is false, $\neg A$ is true. We use **T** and **F** to denote the *truth values* true and false.

Another common truth-functional operation is the *conjunction*: 'and'. The conjunction of sentences A and B will be designated by $A \wedge B$ and has the following truth table:

A	B	$A \wedge B$
T	T	T
F	T	F
T	F	F
F	F	F

$A \wedge B$ is true when and only when both A and B are true. A and B are called the *conjuncts* of $A \wedge B$. Note that there are four rows in the table, corresponding to the number of possible assignments of truth values to A and B.

In natural languages, there are two distinct uses of 'or': the inclusive and the exclusive. According to the inclusive usage, 'A or B' means 'A or B or both', whereas according to the exclusive usage, the meaning is 'A or B, but not both'. We shall introduce a special sign, $\vee$, for the inclusive connective. Its truth table is as follows:

A	B	$A \vee B$
T	T	T
F	T	T
T	F	T
F	F	F

Thus, $A \vee B$ is false when and only when both A and B are false. '$A \vee B$' is called a *disjunction*, with the *disjuncts* A and B.

Another important truth-functional operation is the *conditional*: 'if A, then B'. Ordinary usage is unclear here. Surely, 'if A, then B' is false when the *antecedent* A is true and the *consequent* B is false. However, in other cases, there is no well-defined truth value. For example, the following sentences would be considered neither true nor false:

1. If $1 + 1 = 2$, then Paris is the capital of France.
2. If $1 + 1 \neq 2$, then Paris is the capital of France.
3. If $1 + 1 \neq 2$, then Rome is the capital of France.

Their meaning is unclear, since we are accustomed to the assertion of some sort of relationship (usually causal) between the antecedent and the consequent. We shall make the convention that 'if A, then B' is false when and only when A is true and B is false. Thus, sentences 1–3 are assumed to be true. Let us denote 'if A, then B' by '$A \Rightarrow B$'. An expression '$A \Rightarrow B$' is called a *conditional*. Then $\Rightarrow$ has the following truth table:

A	B	$A \Rightarrow B$
T	T	T
F	T	T
T	F	F
F	F	T

This sharpening of the meaning of 'if A, then B' involves no conflict with ordinary usage, but rather only an extension of that usage.†

A justification of the truth table for $\Rightarrow$ is the fact that we wish 'if A and B, then B' to be true in all cases. Thus, the case in which A and B are true justifies the first line of our truth table for $\Rightarrow$, since (A and B) and B are both true. If A is

†There is a common non-truth-functional interpretation of 'if A, then B' connected with causal laws. The sentence 'if this piece of iron is placed in water at time t, then the iron will dissolve' is regarded as false even in the case that the piece of iron is not placed in water at time t – that is, even when the antecedent is false. Another non-truth-functional usage occurs in so-called counterfactual conditionals, such as 'if Sir Walter Scott had not written any novels, then there would have been no War Between the States'. (This was Mark Twain's contention in *Life on the Mississippi*: 'Sir Walter had so large a hand in making Southern character, as it existed before the war, that he is in great measure responsible for the war'.) This sentence might be asserted to be false even though the antecedent is admittedly false. However, causal laws and counterfactual conditions seem not to be needed in mathematics and logic. For a clear treatment of conditionals and other connectives, see Quine (1951). (The quotation from *Life on the Mississippi* was brought to my attention by Professor J.C. Owings, Jr.)

false and B true, then (A and B) is false while B is true. This corresponds to the second line of the truth table. Finally, if A is false and B is false, (A and B) is false and B is false. This gives the fourth line of the table. Still more support for our definition comes from the meaning of statements such as 'for every x, if x is an odd positive integer, then x^2 is an odd positive integer'. This asserts that, for every x, the statement 'if x is an odd positive integer, then x^2 is an odd positive integer' is true. Now we certainly do not want to consider cases in which x is not an odd positive integer as counterexamples to our general assertion. This supports the second and fourth lines of our truth table. In addition, any case in which x is an odd positive integer and x^2 is an odd positive integer confirms our general assertion. This corresponds to the first line of the table.

Let us denote 'A if and only if B' by '$A \Leftrightarrow B$'. Such an expression is called a *biconditional*. Clearly, $A \Leftrightarrow B$ is true when and only when A and B have the same truth value. Its truth table, therefore is:

A	B	$A \Leftrightarrow B$
T	T	T
F	T	F
T	F	F
F	F	T

The symbols $\neg, \wedge, \vee, \Rightarrow$ and $\Leftrightarrow$ will be called *propositional connectives*.[†] Any sentence built up by application of these connectives has a truth value that depends on the truth values of the constituent sentences. In order to make this dependence apparent, let us apply the name *statement form* to an expression built up from the *statement letters* A, B, C, and so on by appropriate applications of the propositional connectives.

1. All statement letters (capital italic letters) and such letters with numerical subscripts[‡] are statement forms.
2. If $\mathscr{B}$ and $\mathscr{C}$ are statement forms, then so are $(\neg\mathscr{B}), (\mathscr{B} \wedge \mathscr{C})$, $(\mathscr{B} \vee \mathscr{C}), (\mathscr{B} \Rightarrow \mathscr{C})$, and $(\mathscr{B} \Leftrightarrow \mathscr{C})$.
3. Only those expressions are statement forms that are determined to be so by means of conditions 1 and 2.[§]

Some examples of statement forms are $B, (\neg C_2), (D_3 \wedge (\neg B))$, $(((\neg B_1) \vee B_2) \Rightarrow (A_1 \wedge C_2))$, and $(((\neg A) \Leftrightarrow A) \Leftrightarrow (C \Rightarrow (B \vee C)))$.

[†]We have been avoiding and shall in the future avoid the use of quotation marks to form names whenever this is not likely to cause confusion. The given sentence should have quotation marks around each of the connectives. See Quine (1951, pp. 23–27).

[‡]For example, $A_1, A_2, A_{17}, B_{31}, C_2, \ldots$.

[§]This can be rephrased as follows: $\mathscr{C}$ is a statement form if and only if there is a finite sequence $\mathscr{B}_1, \ldots, \mathscr{B}_n (n \geq 1)$ such that $\mathscr{B}_n = \mathscr{C}$ and, if $1 \leqslant i \leqslant n, \mathscr{B}_i$ is either a statement letter or a negation, conjunction, disjunction, conditional or biconditional constructed from previous expressions in the sequence. Notice that we use script letters $\mathscr{A}, \mathscr{B}, \mathscr{C}, \ldots$ to stand for arbitrary expressions, whereas italic letters are used as statement letters.

For every assignment of truth values T or F to the statement letters that occur in a statement form, there corresponds, by virtue of the truth tables for the propositional connectives, a truth value for the statement form. Thus, each statement form determines a *truth function*, which can be graphically represented by a truth table for the statement form. For example, the statement form $(((\neg A) \vee B) \Rightarrow C)$ has the following truth table:

A	B	C	$(\neg A)$	$((\neg A) \vee B)$	$(((\neg A) \vee B) \Rightarrow C)$
T	T	T	F	T	T
F	T	T	T	T	T
T	F	T	F	F	T
F	F	T	T	T	T
T	T	F	F	T	F
F	T	F	T	T	F
T	F	F	F	F	T
F	F	F	T	T	F

Each row represents an assignment of truth values to the statement letters A, B and C and the corresponding truth values assumed by the statement forms that appear in the construction of $(((\neg A) \vee B) \Rightarrow C)$.

The truth table for $((A \Leftrightarrow B) \Rightarrow ((\neg A) \wedge B))$ is as follows:

A	B	$(A \Leftrightarrow B)$	$(\neg A)$	$((\neg A) \wedge B)$	$((A \Leftrightarrow B) \Rightarrow ((\neg A) \wedge B))$
T	T	T	F	F	F
F	T	F	T	T	T
T	F	F	F	F	T
F	F	T	T	F	F

If there are n distinct letters in a statement form, then there are 2^n possible assignments of truth values to the statement letters and, hence, 2^n rows in the truth table.

A truth table can be abbreviated by writing only the full statement form, putting the truth values of the statement letters underneath all occurrences of these letters, and writing, step by step, the truth values of each component statement form under the principal connective of the form[†]. As an example, for $((A \Leftrightarrow B) \Rightarrow ((\neg A) \wedge B))$, we obtain:

$((A$	$\Leftrightarrow$	$B)$	$\Rightarrow$	$((\neg A)$	$\wedge$	$B))$
T	T	T	F	FT	F	T
F	F	T	T	TF	T	T
T	F	F	T	FT	F	F
F	T	F	F	TF	F	F

[†]The *principal connective* of a statement form is the one that is applied last in constructing the form.

Exercises

1.1 Write the truth table for the exclusive usage of 'or'.
1.2 Construct truth tables for the statement forms $((A \Rightarrow B) \vee (\neg A))$ and $((A \Rightarrow (B \Rightarrow C)) \Rightarrow ((A \Rightarrow B) \Rightarrow (A \Rightarrow C)))$.
1.3 Write abbreviated truth tables for $((A \Rightarrow B) \wedge A)$ and $((A \vee (\neg C)) \Leftrightarrow B)$.
1.4 Write the following sentences as statement forms, using statement letters to stand for the *atomic sentences* – that is, those sentences that are not built up out of other sentences.

(a) If Mr Jones is happy, Mrs Jones is not happy, and if Mr Jones is not happy, Mrs Jones is not happy.
(b) Either Sam will come to the party and Max will not, or Sam will not come to the party and Max will enjoy himself.
(c) A sufficient condition for x to be odd is that x is prime.
(d) A necessary condition for a sequence s to converge is that s be bounded.
(e) A necessary and sufficient condition for the sheikh to be happy is that he has wine, women and song.
(f) Fiorello goes to the movies only if a comedy is playing.
(g) The bribe will be paid if and only if the goods are delivered.
(h) If x is positive, x^2 is positive.
(i) Karpov will win the chess tournament unless Kasparov wins today.

1.2 TAUTOLOGIES

A *truth function of n arguments* is defined to be a function of n arguments, the arguments and values of which are the truth values T or F. As we have seen, any statement form containing n distinct statement letters determines a corresponding truth function of n arguments.†

†To be precise, enumerate all statement letters as follows: $A, B, \ldots, Z; A_1, B_1, \ldots, Z_1; A_2, \ldots,$. If a statement form contains the $i_{1\text{th}}, \ldots, i_{n\text{th}}$ statement letters in this enumeration, where $i_1 < \ldots < i_n$, then the corresponding truth function is to have $x_{i_1}, \ldots, x_{i_n}$, in that order, as its arguments, where x_{i_j} corresponds to the $i_{j\text{th}}$ statement letter. For example, $(A \Rightarrow B)$ generates the truth function

x_1	x_2	$f(x_1, x_2)$
T	T	T
F	T	T
T	F	F
F	F	T

whereas $(B \Rightarrow A)$ generates the truth function

x_1	x_2	$g(x_1, x_2)$
T	T	T
F	T	F
T	F	T
F	F	T

A statement form that is always true, no matter what the truth values of its statement letters may be, is called a *tautology*. A statement form is a tautology if and only if its corresponding truth function takes only the value T, or equivalently, if, in its truth table, the column under the statement form contains only *T*s. An example of a tautology is $(A \vee (\neg A))$, the so-called *law of the excluded middle*. Other simple examples are $(\neg(A \wedge (\neg A)))$, $(A \Leftrightarrow (\neg(\neg A)))$, $((A \wedge B) \Rightarrow A)$ and $(A \Rightarrow (A \vee B))$.

$\mathscr{B}$ is said to *logically imply* $\mathscr{C}$ (or, synonymously, $\mathscr{C}$ is a *logical consequence* of $\mathscr{B}$) if and only if every truth assignment to the statement letters of $\mathscr{B}$ and $\mathscr{C}$ that makes $\mathscr{B}$ true also makes $\mathscr{C}$ true. For example, $(A \wedge B)$ logically implies A, A logically implies $(A \vee B)$, and $(A \wedge (A \Rightarrow B))$ logically implies B.

$\mathscr{B}$ and $\mathscr{C}$ are said to be *logically equivalent* if and only if $\mathscr{B}$ and $\mathscr{C}$ receive the same truth value under every assignment of truth values to the statement letters of $\mathscr{B}$ and $\mathscr{C}$. For example, A and $(\neg(\neg A))$ are logically equivalent, as are $(A \wedge B)$ and $(B \wedge A)$.

PROPOSITION 1.1

(a) $\mathscr{B}$ logically implies $\mathscr{C}$ if and only if $(\mathscr{B} \Rightarrow \mathscr{C})$ is a tautology.
(b) $\mathscr{B}$ and $\mathscr{C}$ are logically equivalent if and only if $(\mathscr{B} \Leftrightarrow \mathscr{C})$ is a tautology.

Proof

(a) (i) Assume $\mathscr{B}$ logically implies $\mathscr{C}$. Hence, every truth assignment that makes $\mathscr{B}$ true also makes $\mathscr{C}$ true. Thus, no truth asssignment makes $\mathscr{B}$ true and $\mathscr{C}$ false. Therefore, no truth assignment makes $(\mathscr{B} \Rightarrow \mathscr{C})$ false, that is, every truth assignment makes $(\mathscr{B} \Rightarrow \mathscr{C})$ true. In other words, $(\mathscr{B} \Rightarrow \mathscr{C})$ is a tautology. (ii) Assume $(\mathscr{B} \Rightarrow \mathscr{C})$ is a tautology. Then, for every truth assignment, $(\mathscr{B} \Rightarrow \mathscr{C})$ is true, and, therefore, it is not the case that $\mathscr{B}$ is true and $\mathscr{C}$ false. Hence, every truth assignment that makes $\mathscr{B}$ true makes $\mathscr{C}$ true, that is, $\mathscr{B}$ logically implies $\mathscr{C}$.

(b) $(\mathscr{B} \Leftrightarrow \mathscr{C})$ is a tautology if and only if every truth assignment makes $(\mathscr{B} \Leftrightarrow \mathscr{C})$ true, which is equivalent to saying that every truth assignment gives $\mathscr{B}$ and $\mathscr{C}$ the same truth value, that is, $\mathscr{B}$ and $\mathscr{C}$ are logically equivalent.

By means of a truth table, we have an effective procedure for determining whether a statement form is a tautology. Hence, by Proposition 1.1, we have effective procedures for determining whether a given statement form logically implies another given statement form and whether two given statement forms are logically equivalent.

To see whether a statement form is a tautology, there is another method that is often shorter than the construction of a truth table.

Examples

1. Determine whether $((A \Leftrightarrow ((\neg B) \vee C)) \Rightarrow ((\neg A) \Rightarrow B))$ is a tautology.

Assume that the statement form sometimes is F (line 1). Then $(A \Leftrightarrow ((\neg B) \vee C))$ is T and $((\neg A) \Rightarrow B)$ is F (line 2). Since $((\neg A) \Rightarrow B)$ is F, $(\neg A)$ is T and B is F (line 3). Since $(\neg A)$ is T, A is F (line 4). Since A is F and $(A \Leftrightarrow ((\neg B) \vee C))$ is T, $((\neg B) \vee C)$ is F (line 5). Since $((\neg B) \vee C)$ is F, $(\neg B)$ and C are F (line 6). Since $(\neg B)$ is F, B is T (line 7). But B is both T and F (lines 7 and 3). Hence, it is impossible for the form to be false.

$((A \Leftrightarrow ((\neg B) \vee C)) \Rightarrow ((\neg A) \Rightarrow B))$	
F	1
T F	2
T F	3
F F	4
F	5
F F	6
T	7

2. Determine whether $((A \Rightarrow (B \vee C)) \vee (A \Rightarrow B))$ is a tautology.

Assume that the form is F (line 1). Then $(A \Rightarrow (B \vee C))$ and $(A \Rightarrow B)$ are F (line 2). Since $(A \Rightarrow B)$ is F, A is T and B is F (line 3). Since $(A \Rightarrow (B \vee C))$ is F, A is T and $(B \vee C)$ is F (line 4). Since $(B \vee C)$ is F, B and C are F (line 5). Thus, when A is T, B is F, and C is F, the form is F. Therefore, it is not a tautology.

$((A \Rightarrow (B \vee C)) \vee (A \Rightarrow B))$	
F	1
F F	2
T F	3
T F	4
F F	5

Exercises

1.5 Determine whether the following are tautologies.

(a) $(((A \Rightarrow B) \Rightarrow B) \Rightarrow B)$
(b) $(((A \Rightarrow B) \Rightarrow B) \Rightarrow A)$
(c) $(((A \Rightarrow B) \Rightarrow A) \Rightarrow A)$
(d) $(((B \Rightarrow C) \Rightarrow (A \Rightarrow B)) \Rightarrow (A \Rightarrow B))$
(e) $((A \vee (\neg(B \wedge C))) \Rightarrow ((A \Leftrightarrow C) \vee B))$
(f) $(A \Rightarrow (B \Rightarrow (B \Rightarrow A)))$
(g) $((A \wedge B) \Rightarrow (A \vee C))$
(h) $((A \Leftrightarrow B) \Leftrightarrow (A \Leftrightarrow (B \Leftrightarrow A)))$
(i) $((A \Rightarrow B) \vee (B \Rightarrow A))$
(j) $((\neg(A \Rightarrow B)) \Rightarrow A)$

1.6 Determine whether the following pairs are logically equivalent.

(a) $((A \Rightarrow B) \Rightarrow A)$ and A
(b) $(A \Leftrightarrow B)$ and $((A \Rightarrow B) \wedge (B \Rightarrow A))$
(c) $((\neg A) \vee B)$ and $((\neg B) \vee A)$
(d) $(\neg(A \Leftrightarrow B))$ and $(A \Leftrightarrow (\neg B))$
(e) $(A \vee (B \Leftrightarrow C))$ and $((A \vee B) \Leftrightarrow (A \vee C))$
(f) $(A \Rightarrow (B \Leftrightarrow C))$ and $((A \Rightarrow B) \Leftrightarrow (A \Rightarrow C))$
(g) $(A \wedge (B \Leftrightarrow C))$ and $((A \wedge B) \Leftrightarrow (A \wedge C))$

1.7 Prove:

(a) $(A \Rightarrow B)$ is logically equivalent to $((\neg A) \vee B)$.
(b) $(A \Rightarrow B)$ is logically equivalent to $(\neg(A \wedge (\neg B)))$.

1.8 Prove that $\mathscr{B}$ is logically equivalent to $\mathscr{C}$ if and only if $\mathscr{B}$ logically implies $\mathscr{C}$ and $\mathscr{C}$ logically implies $\mathscr{B}$.

1.9 Show that $\mathscr{B}$ and $\mathscr{C}$ are logically equivalent if and only if, in their truth tables, the columns under $\mathscr{B}$ and $\mathscr{C}$ are the same.

1.10 Prove that $\mathscr{B}$ and $\mathscr{C}$ are logically equivalent if and only if $(\neg\mathscr{B})$ and $(\neg\mathscr{C})$ are logically equivalent.

1.11 Which of the following statement forms are logically implied by $(A \wedge B)$?

(a) A (d) $((\neg A) \vee B)$ (g) $(A \Rightarrow B)$
(b) B (e) $((\neg B) \Rightarrow A)$ (h) $((\neg B) \Rightarrow (\neg A))$
(c) $(A \vee B)$ (f) $(A \Leftrightarrow B)$ (i) $(A \wedge (\neg B))$

1.12 Repeat Exercise 1.11 with $(A \wedge B)$ replaced by $(A \Rightarrow B)$ and by $(\neg(A \Rightarrow B))$, respectively.

1.13 Repeat Exercise 1.11 with $(A \wedge B)$ replaced by $(A \vee B)$.

1.14 Repeat Exercise 1.11 with $(A \wedge B)$ replaced by $(A \Leftrightarrow B)$ and by $(\neg(A \Leftrightarrow B))$, respectively.

A statement form that is false for all possible truth values of its statement letters is said to be *contradictory*. Its truth table has only Fs in the column under the statement form. One example is $(A \Leftrightarrow (\neg A))$:

A	$(\neg A)$	$(A \Leftrightarrow (\neg A))$
T	F	F
F	T	F

Another is $(A \wedge (\neg A))$.

Notice that a statement form $\mathscr{B}$ is a tautology if and only if $(\neg\mathscr{B})$ is contradictory, and vice versa.

A sentence (in some natural language like English or in a formal theory) † that arises from a tautology by the substitution of sentences for all the statement letters, with occurrences of the same statement letter being replaced by the same sentence, is said to be *logically true* (according to the propositional calculus). Such a sentence may be said to be true by virtue of its truth-functional structure alone. An example is the English sentence, 'If it is raining or it is snowing, and it is not snowing, then it is raining', which arises by substitution from the tautology $(((A \vee B) \wedge (\neg B)) \Rightarrow A)$. A sentence that comes from a contradictory statement form by means of substitution is said to be *logically false* (according to the propositional calculus).

Now let us prove a few general facts about tautologies.

†By a formal theory we mean an artificial language in which the notions of *meaningful expressions, axioms* and *rules of inference* are precisely described (see page 34).

PROPOSITION 1.2

If $\mathscr{B}$ and $(\mathscr{B} \Rightarrow \mathscr{C})$ are tautologies, then so is $\mathscr{C}$.

Proof

Assume that $\mathscr{B}$ and $(\mathscr{B} \Rightarrow \mathscr{C})$ are tautologies. If $\mathscr{C}$ took the value F for some assignment of truth values to the statement letters of $\mathscr{B}$ and $\mathscr{C}$, then, since $\mathscr{B}$ is a tautology, $\mathscr{B}$ would take the value T and, therefore, $(\mathscr{B} \Rightarrow \mathscr{C})$ would have the value F for that assignment. This contradicts the assumption that $(\mathscr{B} \Rightarrow \mathscr{C})$ is a tautology. Hence, $\mathscr{C}$ never takes the value F.

PROPOSITION 1.3

If $\mathscr{T}$ is a tautology containing as statement letters $A_1, A_2, \ldots, A_n$, and $\mathscr{B}$ arises from $\mathscr{T}$ by substituting statement forms $\mathscr{S}_1, \mathscr{S}_2, \ldots, \mathscr{S}_n$ for $A_1, A_2, \ldots, A_n$, respectively, then $\mathscr{B}$ is a tautology; that is, substitution in a tautology yields a tautology.

Example
Let $\mathscr{T}$ be $((A_1 \wedge A_2) \Rightarrow A_1)$, let $\mathscr{S}_1$ be $(B \vee C)$ and let $\mathscr{S}_2$ be $(C \wedge D)$. Then $\mathscr{B}$ is $(((B \vee C) \wedge (C \wedge D)) \Rightarrow (B \vee C))$.

Proof

Assume that $\mathscr{T}$ is a tautology. For any assignment of truth values to the statement letters in $\mathscr{B}$, the forms $\mathscr{S}_1, \ldots, \mathscr{S}_n$ have truth values $x_1, \ldots, x_n$ (where each x_n is T or F). If we assign the values $x_1, \ldots, x_n$ to $A_1, \ldots, A_n$, respectively, then the resulting truth value of $\mathscr{T}$ is the truth value of $\mathscr{B}$ for the given assignment of truth values. Since $\mathscr{T}$ is a tautology, this truth value must be T. Thus, $\mathscr{B}$ always takes the value T.

PROPOSITION 1.4

If $\mathscr{C}_1$ arises from $\mathscr{B}_1$ by substitution of $\mathscr{C}$ for one or more occurrences of $\mathscr{B}$, then $((\mathscr{B} \Leftrightarrow \mathscr{C}) \Rightarrow (\mathscr{B}_1 \Leftrightarrow \mathscr{C}_1))$ is a tautology. Hence, if $\mathscr{B}$ and $\mathscr{C}$ are logically equivalent, then so are $\mathscr{B}_1$ and $\mathscr{C}_1$.

Example
Let $\mathscr{B}_1$ be $(C \vee D)$, let $\mathscr{B}$ be C, and let $\mathscr{C}$ be $(\neg(\neg C))$. Then $\mathscr{C}_1$ is $((\neg(\neg C)) \vee D)$. Since C and $(\neg(\neg C))$ are logically equivalent, $(C \vee D)$ and $((\neg(\neg C)) \vee D)$ are also logically equivalent.

Proof

Consider any assignment of truth values to the statement letters. If $\mathscr{B}$ and $\mathscr{C}$ have opposite truth values under this assignment, then $(\mathscr{B} \Leftrightarrow \mathscr{C})$ takes the value F, and, hence, $((\mathscr{B} \Leftrightarrow \mathscr{C}) \Rightarrow (\mathscr{B}_1 \Leftrightarrow \mathscr{C}_1))$ is T. If $\mathscr{B}$ and $\mathscr{C}$ take the same truth values, then so do $\mathscr{B}_1$ and $\mathscr{C}_1$, since $\mathscr{C}_1$ differs from $\mathscr{B}_1$ only in containing $\mathscr{C}$ in some places where $\mathscr{B}_1$ contains $\mathscr{B}$. Therefore, in this case, $(\mathscr{B} \Leftrightarrow \mathscr{C})$ is T, $(\mathscr{B}_1 \Leftrightarrow \mathscr{C}_1)$ is T, and, thus, $((\mathscr{B} \Leftrightarrow \mathscr{C}) \Rightarrow (\mathscr{B}_1 \Leftrightarrow \mathscr{C}_1))$ is T.

Parentheses

It is profitable at this point to agree on some conventions to avoid the use of so many parentheses in writing formulas. This will make the reading of complicated expressions easier.

First, we may omit the outer pair of parentheses of a statement form. (In the case of statement letters, there is no outer pair of parentheses.)

Second, we arbitrarily establish the following decreasing order of strength of the connectives: $\neg, \wedge, \vee, \Rightarrow, \Leftrightarrow$. Now we shall explain a step-by-step process for restoring parentheses to an expression obtained by eliminating some or all parentheses from a statement form. Find the leftmost occurrence of the strongest connective that has not yet been processed.

(i) If the connective is $\neg$ and it precedes a statement form $\mathscr{B}$, restore left and right parentheses to obtain $(\neg\mathscr{B})$.

(ii) If the connective is a binary connective C and it is preceded by a statement form $\mathscr{B}$ and followed by a statement form $\mathscr{D}$, restore left and right parentheses to obtain $(\mathscr{B}\ C\ \mathscr{D})$.

(iii) If neither (i) nor (ii) holds, ignore the connective temporarily and find the leftmost occurrence of the strongest of the remaining unprocessed connectives and repeat (i)–(iii) for that connective.

Examples

Parentheses are restored to the expression in the first line of each of the following in the steps shown:

1. $A \Leftrightarrow (\neg B) \vee C \Rightarrow A$
   $A \Leftrightarrow ((\neg B) \vee C) \Rightarrow A$
   $A \Leftrightarrow (((\neg B) \vee C) \Rightarrow A)$
   $(A \Leftrightarrow (((\neg B) \vee C) \Rightarrow A))$
2. $A \Rightarrow \neg B \Rightarrow C$
   $A \Rightarrow (\neg B) \Rightarrow C$
   $(A \Rightarrow (\neg B)) \Rightarrow C$
   $((A \Rightarrow (\neg B)) \Rightarrow C)$
3. $B \Rightarrow \neg\neg A$
   $B \Rightarrow \neg(\neg A)$

$B \Rightarrow (\neg(\neg A))$
$(B \Rightarrow (\neg(\neg A)))$

4. $A \vee \neg(B \Rightarrow A \vee B)$
$A \vee \neg(B \Rightarrow (A \vee B))$
$A \vee (\neg(B \Rightarrow (A \vee B)))$
$(A \vee (\neg(B \Rightarrow (A \vee B))))$

Not every form can be represented without the use of parentheses. For example, parentheses cannot be further eliminated from $A \Rightarrow (B \Rightarrow C)$, since $A \Rightarrow B \Rightarrow C$ stands for $((A \Rightarrow B) \Rightarrow C)$. Likewise, the remaining parentheses cannot be removed from $\neg(A \vee B)$ or from $A \wedge (B \Rightarrow C)$.

Exercises

1.15 Eliminate as many parentheses as possible from the following forms.

(a) $((B \Rightarrow (\neg A)) \wedge C)$
(b) $(A \vee (B \vee C))$
(c) $(((A \wedge (\neg B)) \wedge C) \vee D)$
(d) $((B \vee (\neg C)) \vee (A \wedge B))$
(e) $((A \Leftrightarrow B) \Leftrightarrow (\neg(C \vee D)))$
(f) $((\neg(\neg(\neg(\mathrm{B} \vee \mathrm{C})))) \Leftrightarrow (\mathrm{B} \Leftrightarrow \mathrm{C}))$
(g) $(\neg((\neg(\neg(B \vee C))) \Leftrightarrow (B \Leftrightarrow C)))$
(h) $((((A \Rightarrow B) \Rightarrow (C \Rightarrow D)) \wedge (\neg A)) \vee C)$

1.16 Restore parentheses to the following forms.

(a) $C \vee \neg A \wedge B$
(b) $B \Rightarrow \neg\neg\neg A \wedge C$
(c) $C \Rightarrow \neg(A \wedge B \Rightarrow C) \wedge A \Leftrightarrow B$
(d) $C \Rightarrow A \Rightarrow A \Leftrightarrow \neg A \vee B$

1.17 Determine whether the following expressions are abbreviations of statement forms and, if so, restore all parentheses.

(a) $\neg\neg A \Leftrightarrow A \Leftrightarrow B \vee C$
(b) $\neg(\neg A \Leftrightarrow A) \Leftrightarrow B \vee C$
(c) $\neg(A \Rightarrow B) \vee C \vee D \Rightarrow B$
(d) $A \Leftrightarrow (\neg A \vee B) \Rightarrow (A \wedge (B \vee C)))$
(e) $\neg A \vee B \vee C \wedge D \Leftrightarrow A \wedge \neg A$
(f) $((A \Rightarrow B \wedge (C \vee D) \wedge (A \vee D))$

1.18 If we write $\neg\mathscr{B}$ instead of $(\neg\mathscr{B})$, $\Rightarrow\mathscr{B}\mathscr{C}$ instead of $(\mathscr{B} \Rightarrow \mathscr{C})$, $\wedge\mathscr{B}\mathscr{C}$ instead of $(\mathscr{B} \wedge \mathscr{C})$, $\vee\mathscr{B}\mathscr{C}$ instead of $(\mathscr{B} \vee \mathscr{C})$, and $\Leftrightarrow\mathscr{B}\mathscr{C}$ instead of $(\mathscr{B} \Leftrightarrow \mathscr{C})$, then there is no need for parentheses. For example, $((\neg A) \wedge (B \Rightarrow (\neg D)))$, which is ordinarily abbreviated as $\neg A \wedge (B \Rightarrow \neg D)$, becomes $\wedge\neg A \Rightarrow B\neg D$. This way of writing forms is called *Polish notation.*

(a) Write $((C \Rightarrow (\neg A)) \vee B)$ and $(C \vee ((B \wedge (\neg D)) \Rightarrow C))$ in this notation.
(b) If we count $\Rightarrow, \wedge, \vee$, and $\Leftrightarrow$ each as $+1$, each statement letter as -1 and $\neg$ as 0, prove that an expression $\mathscr{B}$ in this parenthesis-free notation is a statement form if and only if (i) the sum of the symbols of $\mathscr{B}$ is -1 and (ii) the sum of the symbols in any proper initial segment of $\mathscr{B}$ is non-negative. (If an expression $\mathscr{B}$ can be written in the form $\mathscr{C}\mathscr{D}$, where $\mathscr{C} \neq \mathscr{B}$, then $\mathscr{C}$ is called a *proper initial segment* of $\mathscr{B}$.)
(c) Write the statement forms of Exercise 1.15 in Polish notation.

(d) Determine whether the following expressions are statement forms in Polish notation. If so, write the statement forms in the standard way.

(i) $\neg \Rightarrow ABC \vee AB\neg C$
(ii) $\Rightarrow\Rightarrow AB \Rightarrow\Rightarrow BC \Rightarrow \neg AC$
(iii) $\vee \wedge \vee\neg A\neg BC \wedge \vee AC \vee \neg C\neg A$
(iv) $\vee \wedge B \wedge BBB$

1.19 Determine whether each of the following is a tautology, is contradictory, or neither.

(a) $B \Leftrightarrow (B \vee B)$
(b) $((A \Rightarrow B) \wedge B) \Rightarrow A$
(c) $(\neg A) \Rightarrow (A \wedge B)$
(d) $(A \Rightarrow B) \Rightarrow ((B \Rightarrow C) \Rightarrow (A \Rightarrow C))$
(e) $(A \Leftrightarrow \neg B) \Rightarrow A \vee B$
(f) $A \wedge (\neg(A \vee B))$
(g) $(A \Rightarrow B) \Leftrightarrow ((\neg A) \vee B)$
(h) $(A \Rightarrow B) \Leftrightarrow \neg(A \wedge (\neg B))$
(i) $(B \Leftrightarrow (B \Rightarrow A)) \Rightarrow A$
(j) $A \wedge \neg A \Rightarrow B$

1.20 If A and B are true and C is false, what are the truth values of the following statement forms?

(a) $A \vee C$
(b) $A \wedge C$
(c) $\neg A \wedge \neg C$
(d) $A \Leftrightarrow \neg B \vee C$
(e) $B \vee \neg C \Rightarrow A$
(f) $(B \vee A) \Rightarrow (B \Rightarrow \neg C)$
(g) $(B \Rightarrow \neg A) \Leftrightarrow (A \Leftrightarrow C)$
(h) $(B \Rightarrow A) \Rightarrow ((A \Rightarrow \neg C) \Rightarrow (\neg C \Rightarrow B))$

1.21 If $A \Rightarrow B$ is T, what can be deduced about the truth values of the following?

(a) $A \vee C \Rightarrow B \vee C$
(b) $A \wedge C \Rightarrow B \wedge C$
(c) $\neg A \wedge B \Leftrightarrow A \vee B$

1.22 What further truth values can be deduced from those shown?

(a) $\neg A \vee (A \underset{\text{F}}{\Rightarrow} B)$
(b) $\neg(A \wedge B) \underset{\text{T}}{\Leftrightarrow} \neg A \Rightarrow \neg B$
(c) $(\neg A \vee B) \underset{\text{F}}{\Rightarrow} (A \Rightarrow \neg C)$
(d) $(A \underset{\text{F}}{\Leftrightarrow} B) \underset{\text{T}}{\Leftrightarrow} (C \Rightarrow \neg A)$

1.23 If $A \Leftrightarrow B$ is F, what can be deduced about the truth values of the following?

(a) $A \wedge B$ (b) $A \vee B$ (c) $A \Rightarrow B$ (d) $A \wedge C \Leftrightarrow B \wedge C$

1.24 Repeat Exercise 1.23, but assume that $A \Leftrightarrow B$ is T.

1.25 What further truth values can be deduced from those given?

(a) $(A \wedge B) \underset{\text{F F}}{\Leftrightarrow} (A \vee B)$
(b) $(A \Rightarrow \neg B) \underset{\text{F}}{\Rightarrow} (C \Rightarrow B)$

1.26 (a) Apply Proposition 1.3 when $\mathscr{T}$ is $A_1 \Rightarrow A_1 \vee A_2$, $\mathscr{S}_1$ is $B \wedge D$, and $\mathscr{S}_2$ is $\neg B$.

(b) Apply Proposition 1.4 when $\mathscr{B}_1$ is $(B \Rightarrow C) \wedge D$, $\mathscr{B}$ is $B \Rightarrow C$, and $\mathscr{C}$ is $\neg B \vee C$.

1.27 Show that each statement form in column I is logically equivalent to the form next to it in column II.

	I	*II*	
(a)	$A \Rightarrow (B \Rightarrow C)$	$(A \wedge B) \Rightarrow C$	
(b)	$A \wedge (B \vee C)$	$(A \wedge B) \vee (A \wedge C)$	(Distributive law)
(c)	$A \vee (B \wedge C)$	$(A \vee B) \wedge (A \vee C)$	(Distributive law)
(d)	$(A \wedge B) \vee \neg B$	$A \vee \neg B$	
(e)	$(A \vee B) \wedge \neg B$	$A \wedge \neg B$	
(f)	$A \Rightarrow B$	$\neg B \Rightarrow \neg A$	(Law of the contrapositive)
(g)	$A \Leftrightarrow B$	$B \Leftrightarrow A$	(Biconditional commutativity)
(h)	$(A \Leftrightarrow B) \Leftrightarrow C$	$A \Leftrightarrow (B \Leftrightarrow C)$	(Biconditional associativity)
(i)	$A \Leftrightarrow B$	$(A \wedge B) \vee (\neg A \wedge \neg B)$	
(j)	$\neg(A \Leftrightarrow B)$	$A \Leftrightarrow \neg B$	
(k)	$\neg(A \vee B)$	$(\neg A) \wedge (\neg B)$	(De Morgan's law)
(l)	$\neg(A \wedge B)$	$(\neg A) \vee (\neg B)$	(De Morgan's law)
(m)	$A \vee (A \wedge B)$	A	
(n)	$A \wedge (A \vee B)$	A	
(o)	$A \wedge B$	$B \wedge A$	(Commutativity of conjunction)
(p)	$A \vee B$	$B \vee A$	(Commutativity of disjunction)
(q)	$(A \wedge B) \wedge C$	$A \wedge (B \wedge C)$	(Associativity of conjunction)
(r)	$(A \vee B) \vee C$	$A \vee (B \vee C)$	(Associativity of disjunction)

1.28 Show the logical equivalence of the following pairs.

(a) $\mathscr{T} \wedge \mathscr{B}$ and $\mathscr{B}$, where $\mathscr{T}$ is a tautology.
(b) $\mathscr{T} \vee \mathscr{B}$ and $\mathscr{T}$, where $\mathscr{T}$ is a tautology.
(c) $\mathscr{F} \wedge \mathscr{B}$ and $\mathscr{F}$, where $\mathscr{F}$ is contradictory.
(d) $\mathscr{F} \vee \mathscr{B}$ and $\mathscr{B}$, where $\mathscr{F}$ is contradictory.

1.29

(a) Show the logical equivalence of $\neg(A \Rightarrow B)$ and $A \wedge \neg B$.
(b) Show the logical equivalence of $\neg(A \Leftrightarrow B)$ and $(A \wedge \neg B) \vee (\neg A \wedge B)$.
(c) For each of the following statement forms, find a statement form that is logically equivalent to its negation and in which negation signs apply only to statement letters.

(i) $A \Rightarrow (B \Leftrightarrow \neg C)$
(ii) $\neg A \vee (B \Rightarrow C)$
(iii) $A \wedge (B \vee \neg C)$

1.30 (*Duality*)

(a) If $\mathscr{B}$ is a statement form involving only $\neg$, $\wedge$, and $\vee$, and $\mathscr{B}'$ results from $\mathscr{B}$ by replacing each $\wedge$ by $\vee$ and each $\vee$ by $\wedge$, show that $\mathscr{B}$ is a tautology if and only if $\neg\mathscr{B}'$ is a tautology. Then prove that, if $\mathscr{B} \Rightarrow \mathscr{C}$ is a tau-

tology, then so is $\mathscr{C}' \Rightarrow \mathscr{B}'$, and if $\mathscr{B} \Leftrightarrow \mathscr{C}$ is a tautology, then so is $\mathscr{B}' \Leftrightarrow \mathscr{C}'$. (Here $\mathscr{C}$ is also assumed to involve only $\neg$, $\wedge$ and $\vee$.)

(b) Among the logical equivalences in Exercise 1.27, derive (c) from (b), (e) from (d), (l) from (k), (p) from (o), and (r) from (q).

(c) If $\mathscr{B}$ is a statement form involving only $\neg$, $\wedge$ and $\vee$, and $\mathscr{B}^*$ results from $\mathscr{B}$ by interchanging $\wedge$ and $\vee$ and replacing every statement letter by its negation, show that $\mathscr{B}^*$ is logically equivalent to $\neg\mathscr{B}$. Find a statement form that is logically equivalent to the negation of $(A \vee B \vee C) \wedge (\neg A \vee \neg B \vee D)$, in which $\neg$ applies only to statement letters.

1.31

(a) Prove that a statement form that contains $\Leftrightarrow$ as its only connective is a tautology if and only if each statement letter occurs an even number of times.

(b) Prove that a statement form that contains $\neg$ and $\Leftrightarrow$ as its only connectives is a tautology if and only if $\neg$ and each statement letter occur an even number of times.

1.32 (Shannon, 1938) An electric circuit containing only on-off switches (when a switch is on, it passes current; otherwise it does not) can be represented by a diagram in which, next to each switch, we put a letter representing a necessary and sufficient condition for the switch to be on (see Figure 1.1). The condition that a current flows through this network can be given by the statement form $(A \wedge B) \vee (C \wedge \neg A)$. A statement form representating the circuit shown in Figure 1.2 is $(A \wedge B) \vee ((C \vee A) \wedge \neg B)$, which is logically equivalent to each of the following forms by virtue of the indicated logical equivalence of Exercise 1.27.

$$((A \wedge B) \vee (C \vee A)) \wedge ((A \wedge B) \vee \neg B) \quad \text{(c)}$$
$$((A \wedge B) \vee (C \vee A)) \wedge (A \vee \neg B) \quad \text{(d)}$$
$$((A \wedge B) \vee (A \vee C)) \wedge (A \vee \neg B) \quad \text{(p)}$$
$$(((A \wedge B) \vee A) \vee C) \wedge (A \vee \neg B) \quad \text{(r)}$$
$$(A \vee C) \wedge (A \vee \neg B) \quad \text{(p), (m)}$$
$$A \vee (C \wedge \neg B) \quad \text{(c)}$$

Hence, the given circuit is equivalent to the simpler circuit shown in Figure 1.3. (Two circuits are said to be *equivalent* if current flows through one if and only if it flows through the other, and one circuit is *simpler* if it contains fewer switches.)

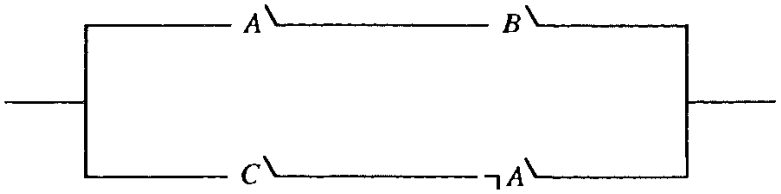


Figure. 1.1

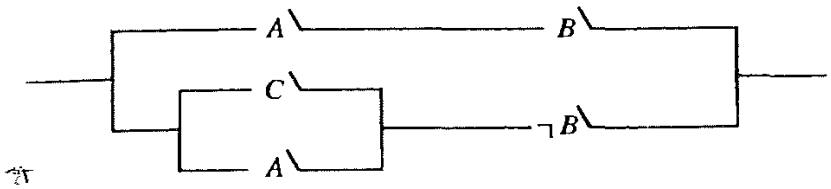


Figure. 1.2

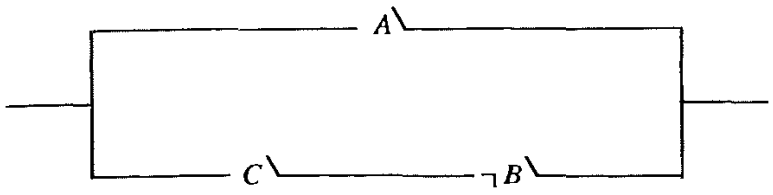


Figure. 1.3

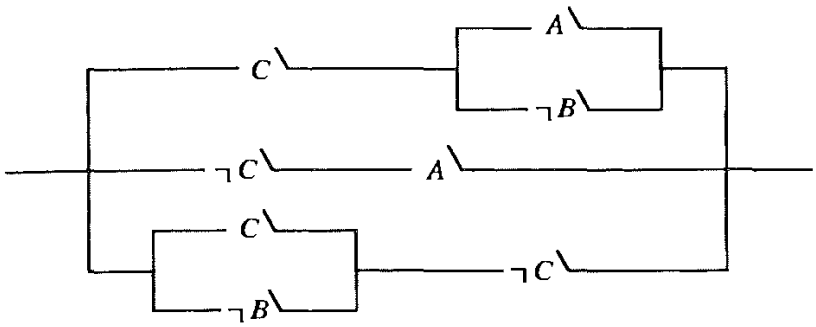


Figure. 1.4

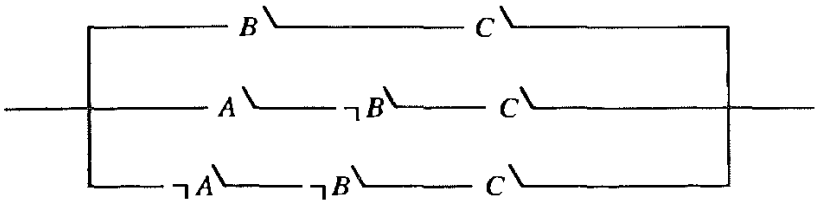


Figure. 1.5

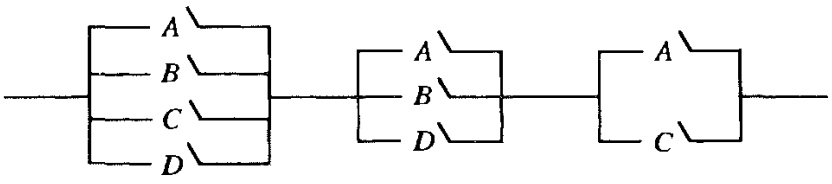


Figure. 1.6

(a) Find simpler equivalent circuits for those shown in Figures 1.4, 1.5 and 1.6.

(b) Assume that each of the three members of a committee votes *yes* on a proposal by pressing a button. Devise as simple a circuit as you can that will allow current to pass when and only when at least two of the members vote in the affirmative.

(c) We wish a light to be controlled by two different wall switches in a room in such a way that flicking either one of these switches will turn the light on if it is off and turn it off if it is on. Construct a simple circuit to do the required job.

1.33 Determine whether the following arguments are logically correct by representing each sentence as a statement form and checking whether the conclusion is logically implied by the conjunction of the assumptions. (To do this, assign T to each assumption and F to the conclusion, and determine whether a contradiction results.)

(a) If Jones is a communist, Jones is an atheist. Jones is an atheist. Therefore, Jones is a communist.
(b) If the temperature and air pressure remained constant, there was no rain. The temperature did remain constant. Therefore, if there was rain, then the air pressure did not remain constant.
(c) If Gorton wins the election, then taxes will increase if the deficit will remain high. If Gorton wins the election, the deficit will remain high. Therefore, if Gorton wins the election, taxes will increase.
(d) If the number x ends in 0, it is divisible by 5. x does not end in 0. Hence, x is not divisible by 5.
(e) If the number x ends in 0, it is divisible by 5. x is not divisible by 5. Hence, x does not end in 0.
(f) If $a = 0$ or $b = 0$, then $ab = 0$. But $ab \neq 0$. Hence, $a \neq 0$ and $b \neq 0$.
(g) A sufficient condition for f to be integrable is that g be bounded. A necessary condition for h to be continuous is that f is integrable. Hence, if g is bounded or h is continuous, then f is integrable.
(h) Smith cannot both be a running star and smoke cigarettes. Smith is not a running star. Therefore, Smith smokes cigarettes.
(i) If Jones drove the car, Smith is innocent. If Brown fired the gun, then Smith is not innocent. Hence, if Brown fired the gun, then Jones did not drive the car.

1.34 Which of the following sets of statement forms are satisfiable, in the sense that there is an assignment of truth values to the statement letters that makes all the forms in the set true?

(a) $A \Rightarrow B$
$B \Rightarrow C$
$C \vee D \Leftrightarrow \neg B$
(b) $\neg (\neg B \vee A)$
$A \vee \neg C)$
$B \Rightarrow \neg C$
(c) $D \Rightarrow B$
$A \vee \neg B$
$\neg(D \wedge A)$
D

1.35 Check each of the following sets of statements for consistency by representing the sentences as statement forms and then testing their conjunction to see whether it is contradictory.

(a) Either the witness was intimidated or, if Doherty committed suicide, a note was found. If the witness was intimidated, then Doherty did not commit suicide. If a note was found, then Doherty committed suicide.

(b) The contract is satisfied if and only if the building is completed by 30 November. The building is completed by 30 November if and only if the electrical subcontractor completes his work by 10 November. The bank loses money if and only if the contract is not satisfied. Yet the electrical subcontractor completes his work by 10 November if and only if the bank loses money.

1.3 ADEQUATE SETS OF CONNECTIVES

Every statement form containing n statement letters generates a corresponding truth function of n arguments. The arguments and values of the function are T or F. Logically equivalent forms generate the same truth function. A natural question is whether all truth functions are so generated.

PROPOSITION 1.5

Every truth function is generated by a statement form involving the connectives $\neg$, $\wedge$ and $\vee$.

Proof

(Refer to Examples 1 and 2 below for clarification.) Let $f(x_1, \ldots, x_n)$ be a truth function. Clearly f can be represented by a truth table of 2^n rows, where each row represents some assignment of truth values to the variables $x_1, \ldots, x_n$, followed by the corresponding value of $f(x_1, \ldots, x_n)$. If $1 \leqslant i \leqslant 2^n$, let C_i be the conjunction $U_1^i \wedge U_2^i \wedge \ldots \wedge U_n^i$, where U_j^i is A_j if, in the ith row of the truth table, x_j takes the value T, and U_j^i is $\neg A_j$ if x_j takes the value F in that row. Let D be the disjunction of all those C_is such that f has the value T for the ith row of the truth table. (If there are no such rows, then f always takes the value F, and we let D be $A_1 \wedge \neg A_1$, which satisfies the theorem.) Notice that D involves only $\neg$, $\wedge$ and $\vee$. To see that D has f as its corresponding truth function, let there be given an assignment of truth values to the statement letters $A_1, \ldots, A_n$, and assume that the corrresponding assignment to the variables $x_1, \ldots, x_n$ is row k of the truth table for f. Then C_k has the value T for this assignment, whereas every other C_i has the value F. If f has the value T for row k, then C_k is a disjunct of D. Hence,

D would also have the value T for this assignment. If f has the value F for row k, then C_k is not a disjunct of D and all the disjuncts take the value F for this assignment. Therefore, D would also have the value F. Thus, D generates the truth function f.

Examples

1.

x_1	x_2	$f(x_1,x_2)$
T	T	F
F	T	T
T	F	T
F	F	T

D is $(\neg A_1 \wedge A_2) \vee (A_1 \wedge \neg A_2) \vee (\neg A_1 \wedge \neg A_2)$.

2.

x_1	x_2	x_3	$g(x_1,x_2,x_3)$
T	T	T	T
F	T	T	F
T	F	T	T
F	F	T	T
T	T	F	F
F	T	F	F
T	F	F	F
F	F	F	T

D is $(A_1 \wedge A_2 \wedge A_3) \vee (A_1 \wedge \neg A_2 \wedge A_3) \vee (\neg A_1 \wedge \neg A_2 \wedge A_3)$
$\vee (\neg A_1 \wedge \neg A_2 \wedge \neg A_3)$.

Exercise

1.36 Find statement forms in the connectives $\neg$, $\wedge$ and $\vee$ that have the following truth functions.

x_1	x_2	x_3	$f(x_1,x_2,x_3)$	$g(x_1,x_2,x_3)$	$h(x_1,x_2,x_3)$
T	T	T	T	T	F
F	T	T	T	T	T
T	F	T	T	T	F
F	F	T	F	F	F
T	T	F	F	T	T
F	T	F	F	F	T
T	F	F	F	T	F
F	F	F	T	F	T

COROLLARY 1.6

Every truth function can be generated by a statement form containing as connectives only $\wedge$ and $\neg$, or only $\vee$ and $\neg$, or only $\Rightarrow$ and $\neg$.

Proof

Notice that $\mathscr{B} \vee \mathscr{C}$ is logically equivalent to $\neg(\neg\mathscr{B} \wedge \neg\mathscr{C})$. Hence, by the second part of Proposition 1.4, any statement form in $\wedge$, $\vee$ and $\neg$ is logically equivalent to a statement form in only $\wedge$ and $\neg$ [obtained by replacing all expressions $\mathscr{B} \vee \mathscr{C}$ by $\neg(\neg\mathscr{B} \wedge \neg\mathscr{C})$]. The other parts of the corollary are similar consequences of the following tautologies:

$$\mathscr{B} \wedge \mathscr{C} \Leftrightarrow \neg(\neg\mathscr{B} \vee \neg\mathscr{C})$$
$$\mathscr{B} \vee \mathscr{C} \Leftrightarrow (\neg\mathscr{B} \Rightarrow \mathscr{C})$$
$$\mathscr{B} \wedge \mathscr{C} \Leftrightarrow \neg(\mathscr{B} \Rightarrow \neg\mathscr{C})$$

We have just seen that there are certain pairs of connectives – for example, $\wedge$ and $\neg$ – in terms of which all truth functions are definable. It turns out that there is a single connective, $\downarrow$ (joint denial), that will do the same job. Its truth table is:

A	B	$A \downarrow B$
T	T	F
F	T	F
T	F	F
F	F	T

$A \downarrow B$ is true when and only when neither A nor B is true. Clearly, $\neg A \Leftrightarrow (A \downarrow A)$ and $(A \wedge B) \Leftrightarrow ((A \downarrow A) \downarrow (B \downarrow B))$ are tautologies. Hence, the adequacy of $\downarrow$ for the construction of all truth functions follows from Corollary 1.6.

Another connective, $|$ (alternative denial), is also adequate for this purpose. Its truth table is

A	B	$A \mid B$
T	T	F
F	T	T
T	F	T
F	F	T

$A \mid B$ is true when and only when not both A and B are true. The adequacy of $|$ follows from the tautologies $\neg A \Leftrightarrow (A \mid A)$ and $(A \vee B) \Leftrightarrow ((A \mid A) \mid (B \mid B))$.

PROPOSITION 1.7

The only binary connectives that alone are adequate for the construction of all truth functions are $\downarrow$ and $|$.

Proof

Assume that $h(A, B)$ is an adequate connective. Now, if h(T,T) were T, then any statement form built up using h alone would take the value T when all

its statement letters take the value T. Hence, $\neg A$ would not be definable in terms of h. So, h(T,T) = F. Likewise, h(F,F) = T. Thus, we have the partial truth table

A	B	$h(A,B)$
T	T	F
F	T	
T	F	
F	F	T

If the second and third entries in the last column are F, F or T, T, then h is $\downarrow$ or $|$. If they are F, T, then $h(A,B) \Leftrightarrow \neg B$ is a tautology; and if they are T, F, then $h(A,B) \Leftrightarrow \neg A$ is a tautology. In both cases, h would be definable in terms of $\neg$. But $\neg$ is not adequate by itself because the only truth functions of one variable definable from it are the identity function and negation itself, whereas the truth function that is always T would not be definable.

Exercises

1.37 Prove that each of the pairs $\Rightarrow$, $\vee$ and $\neg$, $\Leftrightarrow$ is not alone adequate to express all truth functions.

1.38

(a) Prove that $A \vee B$ can be expressed in terms of $\Rightarrow$ alone.
(b) Prove that $A \wedge B$ cannot be expressed in terms of $\Rightarrow$ alone.
(c) Prove that $A \Leftrightarrow B$ cannot be expressed in terms of $\Rightarrow$ alone.

1.39 Show that any two of the connectives $\{\wedge, \Rightarrow, \Leftrightarrow\}$ serve to define the remaining one.

1.40 With one variable A, there are four truth functions:

A	$\neg A$	$A \vee \neg A$	$A \wedge \neg A$
T	F	T	F
F	T	T	F

(a) With two variable A and B, how many truth functions are there ?
(b) How many truth functions of n variables are there ?

1.41 Show that the truth function h determined by $(A \vee B) \Rightarrow \neg C$ generates all truth functions.

1.42 By a *literal* we mean a statement letter or a negation of a statement letter. A statement form is said to be in *disjunctive normal form* (dnf) if it is a disjunction consisting of one or more disjuncts, each of which is a conjunction of one or more literals – for example, $(A \wedge B) \vee (\neg A \wedge C)$, $(A \wedge B \wedge \neg A) \vee (C \wedge \neg B) \vee (A \wedge \neg C)$, $A, A \wedge B$, and $A \vee (B \vee C)$. A form is in *conjunctive normal form* (cnf) if it is a conjunction of one or more conjuncts, each of which is a disjunction of one or more literals – for example, $(B \vee C) \wedge (A \vee B)$, $(B \vee \neg C) \wedge (A \vee D)$, $A \wedge (B \vee A) \wedge (\neg B \vee A)$, $A \vee \neg B$, $A \wedge B$,A. Note that our terminology considers a literal to be a (degenerate) conjunction and a (degenerate) disjunction.

(a) The proof of Proposition 1.5 shows that every statement form $\mathscr{B}$ is logically equivalent to one in disjunctive normal form. By applying this result to $\neg\mathscr{B}$, prove that $\mathscr{B}$ is also logically equivalent to a form in conjunctive normal form.

(b) Find logically equivalent dnfs and cnfs for $\neg(A \Rightarrow B) \vee (\neg A \wedge C)$ and $A \Leftrightarrow ((B \wedge \neg A) \vee C)$. [*Hint*: Instead of relying on Proposition 1.5, it is usually easier to use Exercise 1.27(b) and (c).]

(c) A dnf (cnf) is called *full* if no disjunct (conjunct) contains two occurrences of literals with the same letter and if a letter that occurs in one disjunct (conjunct) also occurs in all the others. For example, $(A \wedge \neg A \wedge B) \vee (A \wedge B)$, $(B \wedge B \wedge C) \vee (B \wedge C)$ and $(B \wedge C) \vee B$ are not full, whereas $(A \wedge B \wedge \neg C) \vee (A \wedge B \wedge C) \vee (A \wedge \neg B \wedge \neg C)$ and $(A \wedge \neg B) \vee (B \wedge A)$ are full dnfs.

(i) Find full dnfs and cnfs logically equivalent to $(A \wedge B) \vee \neg A$ and $\neg(A \Rightarrow B) \vee (\neg A \wedge C)$.

(ii) Prove that every non-contradictory (non-tautologous) statement form $\mathscr{B}$ is logically equivalent to a full dnf (cnf) $\mathscr{C}$, and, if $\mathscr{C}$ contains exactly n letters, then $\mathscr{B}$ is a tautology (is contradictory) if and only if $\mathscr{C}$ has 2^n disjuncts (conjuncts).

(d) For each of the following, find a logically equivalent dnf (cnf), and then find a logically equivalent full dnf (cnf),

(i) $(A \vee B) \wedge (\neg B \vee C)$ (iii) $(A \wedge \neg B) \vee (A \wedge C)$

(ii) $\neg A \vee (B \Rightarrow \neg C)$ (iv) $(A \vee B) \Leftrightarrow \neg C$

(e) Construct statement forms in $\neg$ and $\wedge$ (respectively, in $\neg$ and $\vee$ or in $\neg$ and $\Rightarrow$) logically equivalent to the statement forms in (d).

1.43 A statement form is said to be *satisfiable* if it is true for some assignment of truth values to its statement letters. The problem of determining the satisfiability of an arbitrary cnf plays an important role in the theory of computational complexity; it is an example of a so-called *$\mathscr{NP}$-complete* problem (see Garey and Johnson, 1978).

(a) Show that $\mathscr{B}$ is satisfiable if and only if $\neg\mathscr{B}$ is not a tautology.

(b) Determine whether the following are satisfiable:

(i) $(A \vee B) \wedge (\neg A \vee B \vee C) \wedge (\neg A \vee \neg B \vee \neg C)$

(ii) $((A \Rightarrow B) \vee C) \Leftrightarrow (\neg B \wedge (A \vee C))$

(c) Given a disjunction $\mathscr{D}$ of four or more literals: $L_1 \vee L_2 \vee \ldots \vee L_n$, let $C_1, \ldots, C_{n-2}$ be statement letters that do not occur in $\mathscr{D}$, and construct the cnf $\mathscr{E}$:

$$(L_1 \vee L_2 \vee C_1) \wedge (\neg C_1 \vee L_3 \vee C_2) \wedge (\neg C_2 \vee L_4 \vee C_3) \wedge \ldots \wedge (\neg C_{n-3} \vee L_{n-1} \vee C_{n-2}) \wedge (\neg C_{n-2} \vee L_n \vee \neg C_1)$$

Show that any truth assignment satisfying $\mathscr{D}$ can be extended to a truth assignment satisfying $\mathscr{E}$ and, conversely, any truth assignment satisfying

$\mathscr{E}$ is an extension of a truth assignment satisfying $\mathscr{D}$. (This permits the reduction of the problem of satisfying cnfs to the corresponding problem for cnfs with each conjunct containing at most three literals.)

(d) For a disjunction $\mathscr{D}$ of three literals $L_1 \vee L_2 \vee L_3$, show that a form that has the properties of $\mathscr{E}$ in (c) cannot be constructed, with $\mathscr{E}$ a cnf in which each conjunct contains at most two literals (R. Cowen).

1.44 (*Resolution*) Let $\mathscr{B}$ be a cnf and let C be a statement letter. If C is a disjunct of a disjunction $\mathscr{D}_1$ in $\mathscr{B}$ and $\neg C$ is a disjunct of another disjunction $\mathscr{D}_2$ in $\mathscr{B}$, then a non-empty disjunction obtained by eliminating C from $\mathscr{D}_1$ and $\neg C$ from $\mathscr{D}_2$ and forming the disjunction of the remaining literals (dropping repetitions) is said to be obtained from $\mathscr{B}$ by *resolution on* C. For example, if $\mathscr{B}$ is

$$(A \vee \neg C \vee \neg B) \wedge (\neg A \vee D \vee \neg B) \wedge (C \vee D \vee A),$$

the first and third conjuncts yield $A \vee \neg B \vee D$ by resolution on C. In addition, the first and second conjuncts yield $\neg C \vee \neg B \vee D$ by resolution on A, and the second and third conjuncts yield $D \vee \neg B \vee C$ by resolution on A. If we conjoin to $\mathscr{B}$ any new disjunctions obtained by resolution on all variables, and if we apply the same procedure to the new cnf and keep on iterating this operation, the process must eventually stop, and the final result is denoted $\mathscr{R}es(\mathscr{B})$. In the example, $\mathscr{R}es(\mathscr{B})$ is:

$$(A \vee \neg C \vee \neg B) \wedge (\neg A \vee D \vee \neg B) \wedge (C \vee D \vee A) \wedge (\neg C \vee \neg B \vee D)$$
$$\wedge (D \vee \neg B \vee C) \wedge (A \vee \neg B \vee D) \wedge (D \vee \neg B)$$

(Notice that we have not been careful about specifying the order in which conjuncts or disjuncts are written, since any two arrangements will be logically equivalent.)

(a) Find $\mathscr{R}es(\mathscr{B})$ when $\mathscr{B}$ is each of the following:
  (i) $(A \vee \neg B) \wedge B$
  (ii) $(A \vee B \vee C) \wedge (A \vee \neg B \vee C)$
  (iii) $(A \vee C) \wedge (\neg A \vee B) \wedge (A \vee \neg C) \wedge (\neg A \vee \neg B)$

(b) Show that $\mathscr{B}$ logically implies $\mathscr{R}es(\mathscr{B})$.

(c) If $\mathscr{B}$ is a cnf, let $\mathscr{B}_C$ be the cnf obtained from $\mathscr{B}$ by deleting those conjuncts that contain C or $\neg C$. Let $r_C(\mathscr{B})$ be the cnf that is the conjunction of $\mathscr{B}_C$ and all those disjunctions obtained from $\mathscr{B}$ by resolution on C. For example, if $\mathscr{B}$ is the cnf in the example above, then $r_C(\mathscr{B})$ is $(\neg A \vee D \vee \neg B) \wedge (A \vee \neg B \vee D)$. Prove that, if $r_C(\mathscr{B})$ is satisfiable, then so is $\mathscr{B}$. (R. Cowen)

(d) A cnf $\mathscr{B}$ is said to be a *blatant contradiction* if it contains some letter C and its negation $\neg C$ an conjuncts. An example of a blatant contradiction is $(A \vee B) \wedge B \wedge (C \vee D) \wedge \neg B$. Prove that if $\mathscr{B}$ is unsatisfiable, then $\mathscr{R}es(\mathscr{B})$ is a blatant contradiction. [*Hint*: Use induction on the number n of letters that occur in $\mathscr{B}$. In the induction step, use (c).]

(e) Prove that $\mathscr{B}$ is unsatisfiable if and only if $\mathscr{R}es(\mathscr{B})$ is a blatant contradiction.

1.45 Let $\mathscr{B}$ and $\mathscr{D}$ be statement forms such that $\mathscr{B} \Rightarrow \mathscr{D}$ is a tautology.

(a) If $\mathscr{B}$ and $\mathscr{D}$ have no statement letters in common, show that either $\mathscr{B}$ is contradictory or $\mathscr{D}$ is a tautology.

(b) (*Craig's interpolation theorem*) If $\mathscr{B}$ and $\mathscr{D}$ have the statement letters $B_1, \ldots, B_n$ in common, prove that there is a statement form $\mathscr{C}$ having $B_1, \ldots, B_n$ as its only statement letters such that $\mathscr{B} \Rightarrow \mathscr{C}$ and $\mathscr{C} \Rightarrow \mathscr{D}$ are tautologies.

(c) Solve the special case of (b) in which $\mathscr{B}$ is $(B_1 \Rightarrow A) \wedge (A \Rightarrow B_2)$ and $\mathscr{D}$ is $(B_1 \wedge C) \Rightarrow (B_2 \wedge C)$.

1.46

(a) A certain country is inhabited only by *truth-tellers* (people who always tell the truth) and *liars* (people who always lie). Moreover, the inhabitants will respond only to *yes or no* questions. A tourist comes to a fork in a road where one branch leads to the capital and the other does not. There is no sign indicating which branch to take, but there is a native standing at the fork. What yes or no question should the tourist ask in order to determine which branch to take ? [*Hint*: Let A stand for 'You are a truth-teller' and let B stand for 'The left-hand branch leads to the capital'. Construct, by means of a suitable truth table, a statement form involving A and B such that the native's answer to the question as to whether this statement form is true will be *yes* when and only when B is true.]

(b) In a certain country, there are three kinds of people: *workers* (who always tell the truth), *businessmen* (who always lie), and *students* (who sometimes tell the truth and sometimes lie). At a fork in the road, one branch leads to the capital. A worker, a businessman and a student are standing at the side of the road but are not identifiable in any obvious way. By asking two yes or no questions, find out which fork leads to the capital (Each question may be addressed to any of the three.)

More puzzles of this kind may be found in Smullyan (1978, chap. 3; 1985, chaps 2, 4–8).

1.4 AN AXIOM SYSTEM FOR THE PROPOSITIONAL CALCULUS

Truth tables enable us to answer many of the significant questions concerning the truth-functional connectives, such as whether a given statement form is a tautology, is contradictory, or neither, and whether it logically implies or is logically equivalent to some other given statement form. The more complex parts of logic we shall treat later cannot be handled by truth tables or by any other similar effective procedure. Consequently, another

approach, by means of formal axiomatic theories, will have to be tried. Although, as we have seen, the propositional calculus surrenders completely to the truth table method, it will be instructive to illustrate the axiomatic method in this simple branch of logic.

A formal theory $\mathscr{S}$ is defined when the following conditions are satisfied:

1. A countable set of symbols is given as the symbols of $\mathscr{S}$[†]. A finite sequence of symbols of $\mathscr{S}$ is called an *expression* of $\mathscr{S}$.
2. There is a subset of the set of expressions of $\mathscr{S}$ called the set of *well-formed formulas* (wfs) of $\mathscr{S}$. There is usually an effective procedure to determine whether a given expression is a wf.
3. There is a set of wfs called the set of *axioms* of $\mathscr{S}$. Most often, one can effectively decide whether a given wf is an axiom; in such a case, $\mathscr{S}$ is called an *axiomatic* theory.
4. There is a finite set $R_1, \ldots, R_n$ of relations among wfs, called *rules of inference*. For each R_i, there is a unique positive integer j such that, for every set of j wfs and each wf $\mathscr{B}$, one can effectively decide whether the given j wfs are in the relation R_i to $\mathscr{B}$, and, if so, $\mathscr{B}$ is said to *follow from* or to be a *direct consequence of* the given wfs by virtue of R_i[‡].

A *proof* in $\mathscr{S}$ is a sequence $\mathscr{B}_1, \ldots, \mathscr{B}_k$ of wfs such that, for each i, either $\mathscr{B}_i$ is an axiom of $\mathscr{S}$ or $\mathscr{B}_i$ is a direct consequence of some of the preceding wfs in the sequence by virtue of one of the rules of inference of $\mathscr{S}$.

A *theorem* of $\mathscr{S}$ is a wf $\mathscr{B}$ of $\mathscr{S}$ such that $\mathscr{B}$ is the last wf of some proof in $\mathscr{S}$. Such a proof is called a *proof of* $\mathscr{B}$ *in* $\mathscr{S}$.

Even if $\mathscr{S}$ is axiomatic – that is, if there is an effective procedure for checking any given wf to see whether it is an axiom – the notion of 'theorem' is not necessarily effective since, in general, there is no effective procedure for determining, given any wf $\mathscr{B}$, whether there is a proof of $\mathscr{B}$. A theory for which there is such an effective procedure is said to be *decidable*; otherwise, the theory is said to be *undecidable*.

From an intuitive standpoint, a decidable theory is one for which a machine can be devised to test wfs for theoremhood, whereas, for an undecidable theory, ingenuity is required to determine whether wfs are theorems.

A wf $\mathscr{C}$ is said to be a *consequence* in $\mathscr{S}$ of a set of Γ of wfs if and only if there is a sequence $\mathscr{B}_1, \ldots, \mathscr{B}_k$ of wfs such that $\mathscr{C}$ is $\mathscr{B}_k$ and, for each i, either $\mathscr{B}_i$ is an axiom or $\mathscr{B}_i$ is in Γ, or $\mathscr{B}_i$ is a direct consequence by some rule

[†]These 'symbols' may be thought of as arbitrary objects rather than just linguistic objects. This will become absolutely necessary when we deal with theories with uncountably many symbols in Section 2.12.

[‡]An example of a rule of inference will be the rule *modus ponens* (MP): $\mathscr{C}$ follows from $\mathscr{B}$ and $\mathscr{B} \Rightarrow \mathscr{C}$. According to our precise definition, this rule is the relation consisting of all ordered triples $\langle \mathscr{B}, \mathscr{B} \Rightarrow \mathscr{C}, \mathscr{C} \rangle$, where $\mathscr{B}$ and $\mathscr{C}$ are arbitrary wfs of the formal system.

of inference of some of the preceding wfs in the sequence. Such a sequence is called a *proof* (or *deduction*) *of* $\mathscr{C}$ *from* Γ. The members of Γ are called the *hypotheses* or *premisses* of the proof. We use $\Gamma \vdash \mathscr{C}$ as an abbreviation for '$\mathscr{C}$ is a consequence of Γ'. In order to avoid confusion when dealing with more than one theory, we write $\Gamma \vdash_{\mathscr{S}} \mathscr{C}$, adding the subscript $\mathscr{S}$ to indicate the theory in question.

If Γ is a finite set $\{\mathscr{H}_1, \ldots, \mathscr{H}_m\}$, we write $\mathscr{H}_1, \ldots, \mathscr{H}_m \vdash \mathscr{C}$ instead of $\{\mathscr{H}_1, \ldots, \mathscr{H}_m\} \vdash \mathscr{C}$. If Γ is the empty set $\emptyset$, then $\emptyset \vdash \mathscr{C}$ if and only if $\mathscr{C}$ is a theorem. It is customary to omit the sign '$\emptyset$' and simply write $\vdash \mathscr{C}$. Thus, $\vdash \mathscr{C}$ is another way of asserting that $\mathscr{C}$ is a theorem.

The following are simple properties of the notion of consequence:

1. If $\Gamma \subseteq \Delta$ and $\Gamma \vdash \mathscr{C}$, then $\Delta \vdash \mathscr{C}$.
2. $\Gamma \vdash \mathscr{C}$ if and only if there is a finite subset Δ of Γ such that $\Delta \vdash \mathscr{C}$.
3. If $\Delta \vdash \mathscr{C}$, and for each $\mathscr{B}$ in Δ, $\Gamma \vdash \mathscr{B}$, then $\Gamma \vdash \mathscr{C}$.

Assertion 1 represents the fact that if $\mathscr{C}$ is provable from a set Γ of premisses, then, if we add still more premisses, $\mathscr{C}$ is still provable. Half of 2 follows from 1. The other half is obvious when we notice that any proof of $\mathscr{C}$ from Γ uses only a finite number of premisses from Γ. Proposition 3 is also quite simple: if $\mathscr{C}$ is provable from premisses in Δ, and each premiss in Δ is provable from premisses in Γ, then $\mathscr{C}$ is provable from premisses in Γ.

We now introduce a formal axiomatic theory L for the propositional calculus.

1. The symbols of L are $\neg$, $\Rightarrow$, $(,)$, and the letters A_i with positive integers i as subscripts: $A_1, A_2, A_3, \ldots$. The symbols $\neg$ and $\Rightarrow$ are called *primitive connectives*, and the letters A_i are called *statement letters*.
2. (a) All statement letters are wfs.
   (b) If $\mathscr{B}$ and $\mathscr{C}$ are wfs, then so are $(\neg\mathscr{B})$ and $(\mathscr{B} \Rightarrow \mathscr{C})$.†
   Thus, a wf of L is just a statement form built up from the statement letters A_i by means of the connectives $\neg$ and $\Rightarrow$.
3. If $\mathscr{B}$, $\mathscr{C}$ and $\mathscr{D}$ are wfs of L, then the following are axioms of L:
   (A1) $(\mathscr{B} \Rightarrow (\mathscr{C} \Rightarrow \mathscr{B}))$
   (A2) $((\mathscr{B} \Rightarrow (\mathscr{C} \Rightarrow \mathscr{D})) \Rightarrow ((\mathscr{B} \Rightarrow \mathscr{C}) \Rightarrow (\mathscr{B} \Rightarrow \mathscr{D})))$
   (A3) $(((\neg\mathscr{C}) \Rightarrow (\neg\mathscr{B})) \Rightarrow (((\neg\mathscr{C}) \Rightarrow \mathscr{B}) \Rightarrow \mathscr{C}))$
4. The only rule of inference of L is *modus ponens*: $\mathscr{C}$ is a direct consequence of $\mathscr{B}$ and $(\mathscr{B} \Rightarrow \mathscr{C})$. We shall abbreviate applications of this rule by MP.‡

We shall use our conventions for eliminating parentheses.

†To be precise, we should add the so called extremal clause: (c) An expression is a wf if and only if it can be shown to be a wf on the basis of clauses (a) and (b). This can be made rigorous using as a model the definition in footnote § on page 13.

‡A common English synonym for modus ponens is the *detachment rule*.

Notice that the infinite set of axioms of L is given by means of three axiom schemas (A1)–(A3), with each schema standing for an infinite number of axioms. One can easily check for any given wf whether or not it is an axiom; therefore, L is axiomatic. In setting up the system L, it is our intention to obtain as theorems precisely the class of all tautologies.

We introduce other connectives by definition:

(D1) $(\mathscr{B} \wedge \mathscr{C})$ for $\neg(\mathscr{B} \Rightarrow \neg\mathscr{C})$
(D1) $(\mathscr{B} \vee \mathscr{C})$ for $(\neg\mathscr{B}) \Rightarrow \mathscr{C}$
(D3) $(\mathscr{B} \Leftrightarrow \mathscr{C})$ for $(\mathscr{B} \Rightarrow \mathscr{C}) \wedge (\mathscr{C} \Rightarrow \mathscr{B})$

The meaning of (D1), for example, is that, for any wfs $\mathscr{B}$ and $\mathscr{C}$, '$(\mathscr{B} \wedge \mathscr{C})$' is an abbreviation for '$\neg(\mathscr{B} \Rightarrow \neg\mathscr{C})$'.

LEMMA 1.8 $\vdash_L \mathscr{B} \Rightarrow \mathscr{B}$ for all wfs $\mathscr{B}$.

Proof†

We shall construct a proof in L of $\mathscr{B} \Rightarrow \mathscr{B}$.

1.	$(\mathscr{B} \Rightarrow ((\mathscr{B} \Rightarrow \mathscr{B}) \Rightarrow \mathscr{B})) \Rightarrow ((\mathscr{B} \Rightarrow (\mathscr{B} \Rightarrow \mathscr{B})) \Rightarrow (\mathscr{B} \Rightarrow \mathscr{B}))$	Instance of axiom schema (A2)
2.	$\mathscr{B} \Rightarrow ((\mathscr{B} \Rightarrow \mathscr{B}) \Rightarrow \mathscr{B})$	Axiom schema (A1)
3.	$(\mathscr{B} \Rightarrow (\mathscr{B} \Rightarrow \mathscr{B})) \Rightarrow (\mathscr{B} \Rightarrow \mathscr{B})$	From 1 and 2 by MP
4.	$\mathscr{B} \Rightarrow (\mathscr{B} \Rightarrow \mathscr{B})$	Axiom schema (A1)

†The word 'proof' is used in two distinct senses. First, it has a precise meaning defined above as a certain kind of finite sequence of wfs of L. However, in another sense, it also designates certain sequences of the English language (supplemented by various technical terms) that are supposed to serve as an argument justifying some assertion about the language L (or other formal theories). In general, the language we are studying (in this case, L) is called the *object language*, while the language in which we formulate and prove statements about the object language is called the *metalanguage*. The metalanguage might also be formalized and made the subject of study, which we would carry out in a metametalanguage, and so on. However, we shall use the English language as our (unformalized) metalanguage, although, for a substantial part of this book, we use only a mathematically weak portion of the English language. The contrast between object language and metalanguage is also present in the study of a foreign language; for example, in a Sanskrit class, Sanskrit is the object language, while the metalanguage, the language we use, is English. The distinction between *proof* and *metaproof* (i.e., a proof in the metalanguage) leads to a distinction between theorems of the object language and *metatheorems* of the metalanguage. To avoid confusion, we generally use 'proposition' instead of 'metatheorem'. The word 'metamathematics' refers to the study of logical and mathematical object languages; sometimes the word is restricted to those investigations that use what appear to the metamathematician to be constructive (or so-called finitary) methods.

5. $\mathscr{B} \Rightarrow \mathscr{B}$ From 3 and 4 by MP†

Exercise

1.47 Prove:

(a) $\vdash_{\mathrm{L}} (\neg\mathscr{B} \Rightarrow \mathscr{B}) \Rightarrow \mathscr{B}$

(b) $\mathscr{B} \Rightarrow \mathscr{C}, \mathscr{C} \Rightarrow \mathscr{D} \vdash_{\mathrm{L}} \mathscr{B} \Rightarrow \mathscr{D}$

(c) $\mathscr{B} \Rightarrow (\mathscr{C} \Rightarrow \mathscr{D}) \vdash_{\mathrm{L}} \mathscr{C} \Rightarrow (\mathscr{B} \Rightarrow \mathscr{D})$

(d) $\vdash_{\mathrm{L}} (\neg\mathscr{C} \Rightarrow \neg\mathscr{B}) \Rightarrow (\mathscr{B} \Rightarrow \mathscr{C})$

In mathematical arguments, one often proves a statement $\mathscr{C}$ on the assumption of some other statement $\mathscr{B}$ and then concludes that 'if $\mathscr{B}$, then $\mathscr{C}$' is true. This procedure is justified for the system L by the following theorem.

PROPOSITION 1.9 (DEDUCTION THEOREM)‡

If Γ is a set of wfs and $\mathscr{B}$ and $\mathscr{C}$ are wfs, and $\Gamma, \mathscr{B} \vdash \mathscr{C}$, then $\Gamma \vdash \mathscr{B} \Rightarrow \mathscr{C}$. In particular, if $\mathscr{B} \vdash \mathscr{C}$, then $\vdash \mathscr{B} \Rightarrow \mathscr{C}$ (Herbrand, 1930).

Proof

Let $\mathscr{C}_1, \ldots, \mathscr{C}_n$ be a proof of $\mathscr{C}$ form $\Gamma \cup \{\mathscr{B}\}$, where $\mathscr{C}_n$ is $\mathscr{C}$. Let us prove, by induction on j, that $\Gamma \vdash \mathscr{B} \Rightarrow \mathscr{C}_j$ for $1 \leqslant j \leqslant n$. First of all, $\mathscr{C}_1$ must be either in Γ or an axiom of L or $\mathscr{B}$ itself. By axiom schema (A1), $\mathscr{C}_1 \Rightarrow (\mathscr{B} \Rightarrow \mathscr{C}_1)$ is an axiom. Hence, in the first two cases, by MP, $\Gamma \vdash \mathscr{B} \Rightarrow \mathscr{C}_1$. For the third case, when $\mathscr{C}_1$ is $\mathscr{B}$, we have $\vdash \mathscr{B} \Rightarrow \mathscr{C}_1$ by Lemma 1.8, and, therefore, $\Gamma \vdash \mathscr{B} \Rightarrow \mathscr{C}_1$. This takes care of the case $j = 1$. Assume now that $\Gamma \vdash \mathscr{B} \Rightarrow \mathscr{C}_k$ for all $k < j$. Either $\mathscr{C}_j$ is an axiom, or $\mathscr{C}_j$ is in Γ, or $\mathscr{C}_j$ is $\mathscr{B}$, or $\mathscr{C}_j$ follows by modus ponens from some $\mathscr{C}_\ell$ and $\mathscr{C}_m$, where $\ell < j$, $m < j$, and $\mathscr{C}_m$ has the form $\mathscr{C}_\ell \Rightarrow \mathscr{C}_j$. In the first three cases, $\Gamma \vdash \mathscr{B} \Rightarrow \mathscr{C}_j$ as in the case $j = 1$ above. In the last case, we have, by inductive hypothesis, $\Gamma \vdash \mathscr{B} \Rightarrow \mathscr{C}_\ell$ and $\Gamma \vdash \mathscr{B} \Rightarrow (\mathscr{C}_\ell \Rightarrow \mathscr{C}_j)$. But, by axiom schema (A2), $\vdash (\mathscr{B} \Rightarrow (\mathscr{C}_\ell \Rightarrow \mathscr{C}_j)) \Rightarrow ((\mathscr{B} \Rightarrow \mathscr{C}_\ell) \Rightarrow (\mathscr{B} \Rightarrow \mathscr{C}_j))$. Hence, by MP, $\Gamma \vdash (\mathscr{B} \Rightarrow \mathscr{C}_\ell) \Rightarrow (\mathscr{B} \Rightarrow \mathscr{C}_j)$, and, again by MP, $\Gamma \vdash \mathscr{B} \Rightarrow \mathscr{C}_j$. Thus, the proof by induction is complete. The case $j = n$ is the desired result. [Notice that, given a deduction of $\mathscr{C}$ from Γ and $\mathscr{B}$, the proof just given enables us to construct a deduction of $\mathscr{B} \Rightarrow \mathscr{C}$

†The reader should not be discouraged by the apparently unmotivated step 1 of the proof. As in most proofs, we actually begin with the desired result, $\mathscr{B} \Rightarrow \mathscr{B}$, and then look for an appropriate axiom that may lead by MP to that result. A mixture of ingenuity and experimentation leads to a suitable instance of axiom (A2).

‡For the remainder of the chapter, unless something is said to the contrary, we shall omit the subscript L in $\vdash_L$. In addition, we shall use $\Gamma, \mathscr{B} \vdash \mathscr{C}$ to stand for $\Gamma \cup \{\mathscr{B}\} \vdash \mathscr{C}$. In general, we let $\Gamma, \mathscr{B}_1, \ldots, \mathscr{B}_n \vdash \mathscr{C}$ stand for $\Gamma \cup \{\mathscr{B}_1, \ldots, \mathscr{B}_n\} \vdash \mathscr{C}$.

from Γ. Also note that axiom schema (A3) was not used in proving the deduction theorem.]

COROLLARY 1.10

(a) $\mathscr{B} \Rightarrow \mathscr{C}, \mathscr{C} \Rightarrow \mathscr{D} \vdash \mathscr{B} \Rightarrow \mathscr{D}$
(b) $\mathscr{B} \Rightarrow (\mathscr{C} \Rightarrow \mathscr{D}), \mathscr{C} \vdash \mathscr{B} \Rightarrow \mathscr{D}$

Proof
For part (a):

1. $\mathscr{B} \Rightarrow \mathscr{C}$	Hyp (abbreviation for 'hypothesis')
2. $\mathscr{C} \Rightarrow \mathscr{D}$	Hyp
3. $\mathscr{B}$	Hyp
4. $\mathscr{C}$	1, 3, MP
5. $\mathscr{D}$	2, 4, MP

Thus, $\mathscr{B} \Rightarrow \mathscr{C}, \mathscr{C} \Rightarrow \mathscr{D}, \mathscr{B} \vdash \mathscr{D}$. So, by the deduction theorem, $\mathscr{B} \Rightarrow \mathscr{C}, \mathscr{C} \Rightarrow \mathscr{D} \vdash \mathscr{B} \Rightarrow \mathscr{D}$.

To prove (b), use the deduction theorem.

LEMMA 1.11

For any wfs $\mathscr{B}$ and $\mathscr{C}$, the following wfs are theorems of L.

(a) $\neg\neg\mathscr{B} \Rightarrow \mathscr{B}$
(b) $\mathscr{B} \Rightarrow \neg\neg\mathscr{B}$
(c) $\neg\mathscr{B} \Rightarrow (\mathscr{B} \Rightarrow \mathscr{C})$
(d) $(\neg\mathscr{C} \Rightarrow \neg\mathscr{B}) \Rightarrow (\mathscr{B} \Rightarrow \mathscr{C})$
(e) $(\mathscr{B} \Rightarrow \mathscr{C}) \Rightarrow (\neg\mathscr{C} \Rightarrow \neg\mathscr{B})$
(f) $\mathscr{B} \Rightarrow (\neg\mathscr{C} \Rightarrow \neg(\mathscr{B} \Rightarrow \mathscr{C}))$
(g) $(\mathscr{B} \Rightarrow \mathscr{C}) \Rightarrow ((\neg\mathscr{B} \Rightarrow \mathscr{C}) \Rightarrow \mathscr{C})$

Proof

(a) $\vdash \neg\neg\mathscr{B} \Rightarrow \mathscr{B}$

1. $(\neg\mathscr{B} \Rightarrow \neg\neg\mathscr{B}) \Rightarrow ((\neg\mathscr{B} \Rightarrow \neg\mathscr{B}) \Rightarrow \mathscr{B})$	Axiom (A3)
2. $\neg\mathscr{B} \Rightarrow \neg\mathscr{B}$	Lemma 1.8 †
3. $(\neg\mathscr{B} \Rightarrow \neg\neg\mathscr{B}) \Rightarrow \mathscr{B}$	1, 2, Corollary 1.10(b)
4. $\neg\neg\mathscr{B} \Rightarrow (\neg\mathscr{B} \Rightarrow \neg\neg\mathscr{B})$	Axiom (A1)
5. $\neg\neg\mathscr{B} \Rightarrow \mathscr{B}$	3, 4, Corollary 1.10(a)

†Instead of writing a complete proof of $\neg\mathscr{B} \Rightarrow \neg\mathscr{B}$, we simply cite Lemma 1.8. In this way, we indicate how the proof of $\neg\neg\mathscr{B} \Rightarrow \mathscr{B}$ could be written if we wished to take the time and space to do so. This is, of course, nothing more than the ordinary application of previously proved theorems.

(b) $\vdash \mathscr{B} \Rightarrow \neg\neg\mathscr{B}$

1. $(\neg\neg\neg\mathscr{B} \Rightarrow \neg\mathscr{B}) \Rightarrow ((\neg\neg\neg\mathscr{B} \Rightarrow \mathscr{B}) \Rightarrow \neg\neg\mathscr{B})$	Axiom (A3)
2. $\neg\neg\neg\mathscr{B} \Rightarrow \neg\mathscr{B}$	Part (a)
3. $(\neg\neg\neg\mathscr{B} \Rightarrow \mathscr{B}) \Rightarrow \neg\neg\mathscr{B}$	1, 2, MP
4. $\mathscr{B} \Rightarrow (\neg\neg\neg\mathscr{B} \Rightarrow \mathscr{B})$	Axiom (A1)
5. $\mathscr{B} \Rightarrow \neg\neg\mathscr{B}$	3, 4, Corollary 1.10(a)

(c) $\vdash \neg\mathscr{B} \Rightarrow (\mathscr{B} \Rightarrow \mathscr{C})$

1. $\neg\mathscr{B}$	Hyp
2. $\mathscr{B}$	Hyp
3. $\mathscr{B} \Rightarrow (\neg\mathscr{C} \Rightarrow \mathscr{B})$	Axiom (A1)
4. $\neg\mathscr{B} \Rightarrow (\neg\mathscr{C} \Rightarrow \neg\mathscr{B})$	Axiom (A1)
5. $\neg\mathscr{C} \Rightarrow \mathscr{B}$	2, 3, MP
6. $\neg\mathscr{C} \Rightarrow \neg\mathscr{B}$	1, 4, MP
7. $(\neg\mathscr{C} \Rightarrow \neg\mathscr{B}) \Rightarrow ((\neg\mathscr{C} \Rightarrow \mathscr{B}) \Rightarrow \mathscr{C})$	Axiom (A3)
8. $(\neg\mathscr{C} \Rightarrow \mathscr{B}) \Rightarrow \mathscr{C}$	6, 7, MP
9. $\mathscr{C}$	5, 8, MP
10. $\neg\mathscr{B}, \mathscr{B} \vdash \mathscr{C}$	1–9
11. $\neg\mathscr{B} \vdash \mathscr{B} \Rightarrow \mathscr{C}$	10, Deduction theorem
12. $\vdash \neg\mathscr{B} \Rightarrow (\mathscr{B} \Rightarrow \mathscr{C})$	11, Deduction theorem

(d) $\vdash (\neg\mathscr{C} \Rightarrow \neg\mathscr{B}) \Rightarrow (\mathscr{B} \Rightarrow \mathscr{C})$

1. $\neg\mathscr{C} \Rightarrow \neg\mathscr{B}$	Hyp
2. $(\neg\mathscr{C} \Rightarrow \neg\mathscr{B}) \Rightarrow ((\neg\mathscr{C} \Rightarrow \mathscr{B}) \Rightarrow \mathscr{C})$	Axiom (A3)
3. $\mathscr{B} \Rightarrow (\neg\mathscr{C} \Rightarrow \mathscr{B})$	Axiom (A1)
4. $(\neg\mathscr{C} \Rightarrow \mathscr{B}) \Rightarrow \mathscr{C}$	1, 2, MP
5. $\mathscr{B} \Rightarrow \mathscr{C}$	3, 4, Corollary 1.10(a)
6. $\neg\mathscr{C} \Rightarrow \neg\mathscr{B} \vdash \mathscr{B} \Rightarrow \mathscr{C}$	1–5
7. $\vdash (\neg\mathscr{C} \Rightarrow \neg\mathscr{B}) \Rightarrow (\mathscr{B} \Rightarrow \mathscr{C})$	6, deduction theorem

(e) $\vdash (\mathscr{B} \Rightarrow \mathscr{C}) \Rightarrow (\neg\mathscr{C} \Rightarrow \neg\mathscr{B})$

1. $\mathscr{B} \Rightarrow \mathscr{C}$	Hyp
2. $\neg\neg\mathscr{B} \Rightarrow \mathscr{B}$	Part (a)
3. $\neg\neg\mathscr{B} \Rightarrow \mathscr{C}$	1, 2, Corollary 1.10(a)
4. $\mathscr{C} \Rightarrow \neg\neg\mathscr{C}$	Part (b)
5. $\neg\neg\mathscr{B} \Rightarrow \neg\neg\mathscr{C}$	3, 4, Corollary 1.10(a)
6. $(\neg\neg\mathscr{B} \Rightarrow \neg\neg\mathscr{C}) \Rightarrow (\neg\mathscr{C} \Rightarrow \neg\mathscr{B})$	Part (d)
7. $\neg\mathscr{C} \Rightarrow \neg\mathscr{B}$	5, 6, MP
8. $\mathscr{B} \Rightarrow \mathscr{C} \vdash \neg\mathscr{C} \Rightarrow \neg\mathscr{B}$	1–7
9. $\vdash (\mathscr{B} \Rightarrow \mathscr{C}) \Rightarrow (\neg\mathscr{C} \Rightarrow \neg\mathscr{B})$	8, deduction theorem

(f) $\vdash \mathscr{B} \Rightarrow (\neg\mathscr{C} \Rightarrow \neg(\mathscr{B} \Rightarrow \mathscr{C}))$.

Clearly, $\mathscr{B}, \mathscr{B} \Rightarrow \mathscr{C} \vdash \mathscr{C}$ by MP. Hence, $\vdash \mathscr{B} \Rightarrow ((\mathscr{B} \Rightarrow \mathscr{C}) \Rightarrow \mathscr{C}))$ by two uses of the deduction theorem. Now, by (e), $\vdash ((\mathscr{B} \Rightarrow \mathscr{C}) \Rightarrow \mathscr{C}) \Rightarrow (\neg\mathscr{C} \Rightarrow \neg(\mathscr{B} \Rightarrow \mathscr{C}))$. Hence, by Corollary 1.10(a), $\vdash \mathscr{B} \Rightarrow (\neg\mathscr{C} \Rightarrow \neg(\mathscr{B} \Rightarrow \mathscr{C}))$.

(g) $\vdash (\mathscr{B} \Rightarrow \mathscr{C}) \Rightarrow ((\neg\mathscr{B} \Rightarrow \mathscr{C}) \Rightarrow \mathscr{C})$

1. $\mathscr{B} \Rightarrow \mathscr{C}$	Hyp
2. $\neg\mathscr{B} \Rightarrow \mathscr{C}$	Hyp
3. $(\mathscr{B} \Rightarrow \mathscr{C}) \Rightarrow (\neg\mathscr{C} \Rightarrow \neg\mathscr{B})$	Part (e)
4. $\neg\mathscr{C} \Rightarrow \neg\mathscr{B}$	1, 3, MP
5. $(\neg\mathscr{B} \Rightarrow \mathscr{C}) \Rightarrow (\neg\mathscr{C} \Rightarrow \neg\neg\mathscr{B})$	Part (e)
6. $\neg\mathscr{C} \Rightarrow \neg\neg\mathscr{B}$	2, 5, MP
7. $(\neg\mathscr{C} \Rightarrow \neg\neg\mathscr{B}) \Rightarrow ((\neg\mathscr{C} \Rightarrow \neg\mathscr{B}) \Rightarrow \mathscr{C})$	Axiom (A3)
8. $(\neg\mathscr{C} \Rightarrow \neg\mathscr{B}) \Rightarrow \mathscr{C}$	6, 7, MP
9. $\mathscr{C}$	4, 8, MP
10. $\mathscr{B} \Rightarrow \mathscr{C}, \neg\mathscr{B} \Rightarrow \mathscr{C} \vdash \mathscr{C}$	1–9
11. $\mathscr{B} \Rightarrow \mathscr{C} \vdash (\neg\mathscr{B} \Rightarrow \mathscr{C}) \Rightarrow \mathscr{C}$	10, deduction theorem
12. $\vdash (\mathscr{B} \Rightarrow \mathscr{C}) \Rightarrow ((\neg\mathscr{B} \Rightarrow \mathscr{C}) \Rightarrow \mathscr{C})$	11, deduction theorem

Exercises

1.48 Show that the following wfs are theorems of L.

(a) $\mathscr{B} \Rightarrow (\mathscr{B} \vee \mathscr{C})$
(b) $\mathscr{B} \Rightarrow (\mathscr{C} \vee \mathscr{B})$
(c) $\mathscr{C} \vee \mathscr{B} \Rightarrow \mathscr{B} \vee \mathscr{C}$
(d) $\mathscr{B} \wedge \mathscr{C} \Rightarrow \mathscr{B}$
(e) $\mathscr{B} \wedge \mathscr{C} \Rightarrow \mathscr{C}$
(f) $(\mathscr{B} \Rightarrow \mathscr{D}) \Rightarrow ((\mathscr{C} \Rightarrow \mathscr{D}) \Rightarrow (\mathscr{B} \vee \mathscr{C} \Rightarrow \mathscr{D}))$
(g) $((\mathscr{B} \Rightarrow \mathscr{C}) \Rightarrow \mathscr{B}) \Rightarrow \mathscr{B}$
(h) $\mathscr{B} \Rightarrow (\mathscr{C} \Rightarrow (\mathscr{B} \wedge \mathscr{C}))$

1.49 Exhibit a complete proof in L of Lemma 1.11(c). [*Hint:* Apply the procedure used in the proof of the deduction theorem to the demonstration given earlier of Lemma 1.11(c).] Greater fondness for the deduction theorem will result if the reader tries to prove all of Lemma 1.11 without using the deduction theorem.

It is our purpose to show that a wf of L is a theorem of L if and only if it is a tautology. Half of this is very easy.

PROPOSITION 1.12

Every theorem of L is a tautology.

Proof

As an exercise, verify that all the axioms of L are tautologies. By Proposition 1.2, modus ponens leads from tautologies to other tautologies. Hence, every theorem of L is a tautology.

The following lemma is to be used in the proof that every tautology is a theorem of L.

LEMMA 1.13

Let $\mathscr{B}$ be a wf and let $B_1, \ldots, B_k$ be the statement letters that occur in $\mathscr{B}$. For a given assignment of truth values to $B_1, \ldots, B_k$, let B_j' be B_j if B_j takes the value T; and let B_j' be $\neg B_j$ if B_j takes the value F. Let $\mathscr{B}'$ be $\mathscr{B}$ if $\mathscr{B}$ takes the value T under the assignment, and let $\mathscr{B}'$ be $\neg\mathscr{B}$ if $\mathscr{B}$ takes the value F. Then $B_1', \ldots, B_k' \vdash \mathscr{B}'$.

For example, let $\mathscr{B}$ be $\neg(\neg A_2 \Rightarrow A_5)$. Then for each row of the truth table

A_2	A_5	$\neg(\neg A_2 \Rightarrow A_5)$
T	T	F
F	T	F
T	F	F
F	F	T

Lemma 1.13 asserts a corresponding deducibility relation. For instance, corresponding to the third row there is $A_2, \neg A_5 \vdash \neg\neg(\neg A_2 \Rightarrow A_5)$, and to the fourth row, $\neg A_2, \neg A_5 \vdash \neg(\neg A_2 \Rightarrow A_5)$.

Proof

The proof is by induction on the number n of occurrences of $\neg$ and $\Rightarrow$ in $\mathscr{B}$. (We assume $\mathscr{B}$ written without abbreviations.) If $n = 0$, $\mathscr{B}$ is just a statement letter B_1, and then the lemma reduces to $B_1 \vdash B_1$ and $\neg B_1 \vdash \neg B_1$. Assume now that the lemma holds for all $j < n$.

Case 1. $\mathscr{B}$ is $\neg\mathscr{C}$. Then $\mathscr{C}$ has fewer than n occurrences of $\neg$ and $\Rightarrow$.

Subcase 1a. Let $\mathscr{C}$ take the value T under the given truth value assignment. Then $\mathscr{B}$ takes the value F. So, $\mathscr{C}'$ is $\mathscr{C}$ and $\mathscr{B}'$ is $\neg\mathscr{B}$. By the inductive hypothesis applied to $\mathscr{C}$, we have $B_1', \ldots, B_k' \vdash \mathscr{C}$. Then, by Lemma 1.11(b) and MP, $B_1', \ldots, B_k' \vdash \neg\neg\mathscr{C}$. But $\neg\neg\mathscr{C}$ is $\mathscr{B}'$.

Subcase 1b. Let $\mathscr{C}$ take the value F. Then $\mathscr{B}$ takes the value T. So, $\mathscr{C}'$ is $\neg\mathscr{C}$ and $\mathscr{B}'$ is $\mathscr{B}$, By inductive hypothesis, $B_1', \ldots, B_k' \vdash \neg\mathscr{C}$. But $\neg\mathscr{C}$ is $\mathscr{B}'$.

Case 2. $\mathscr{B}$ is $\mathscr{C} \Rightarrow \mathscr{D}$. Then $\mathscr{C}$ and $\mathscr{D}$ have fewer occurrences of $\neg$ and $\Rightarrow$ than $\mathscr{B}$. So, by inductive hypothesis, $B_1', \ldots, B_k' \vdash \mathscr{C}'$ and $B_1', \ldots, B_k' \vdash \mathscr{D}'$.

Subcase 2a. $\mathscr{C}$ takes the value F. Then $\mathscr{B}$ takes the value T. So, $\mathscr{C}'$ is $\neg\mathscr{C}$ and $\mathscr{B}'$ is $\mathscr{B}$. Hence, $B_1', \ldots, B_k' \vdash \neg\mathscr{C}$. By Lemma 1.11(c) and MP, $B_1', \ldots, B_k' \vdash \mathscr{C} \Rightarrow \mathscr{D}$. But $\mathscr{C} \Rightarrow \mathscr{D}$ is $\mathscr{B}'$.

Subcase 2b. $\mathscr{D}$ takes the value T. Then $\mathscr{B}$ takes the value T. So, $\mathscr{D}'$ is $\mathscr{D}$ and $\mathscr{B}'$ is $\mathscr{B}$. Hence, $B_1', \ldots, B_k' \vdash \mathscr{D}$. Then, by axiom (A1) and MP, $B_1', \ldots, B_k' \vdash \mathscr{C} \Rightarrow \mathscr{D}$. But $\mathscr{C} \Rightarrow \mathscr{D}$ is $\mathscr{B}'$.

Subcase 2c. $\mathscr{C}$ takes the value T and $\mathscr{D}$ takes the value F. Then $\mathscr{B}$ takes the value F. So, $\mathscr{C}'$ is $\mathscr{C}$, $\mathscr{D}'$ is $\neg\mathscr{D}$, and $\mathscr{B}'$ is $\neg\mathscr{B}$. Therefore, $B_1', \ldots, B_k' \vdash \mathscr{C}$ and $B_1', \ldots, B_k' \vdash \neg\mathscr{D}$. Hence, by Lemma 1.11(f) and MP, $B_1', \ldots, B_k' \vdash \neg(\mathscr{C} \Rightarrow \mathscr{D})$. But $\neg(\mathscr{C} \Rightarrow \mathscr{D})$ is $\mathscr{B}'$.

PROPOSITION 1.14 (COMPLETENESS THEOREM)

If a wf $\mathscr{B}$ of L is a tautology, then it is a theorem of L.

Proof

(Kalmár, 1935) Assume $\mathscr{B}$ is a tautology, and let $B_1, \ldots, B_k$ be the statement letters in $\mathscr{B}$. For any truth value assignment to $B_1, \ldots, B_k$, we have, by Lemma 1.13, $B'_1, \ldots, B'_k \vdash \mathscr{B}$. ($\mathscr{B}'$ is $\mathscr{B}$ because $\mathscr{B}$ always takes the value T.) Hence, when B'_k is given the value T, we obtain $B'_1, \ldots, B'_{k-1}, B_k \vdash \mathscr{B}$, and, when B_k is given the value F, we obtain $B'_1, \ldots, B'_{k-1}, \neg B_k \vdash \mathscr{B}$. So, by the deduction theorem, $B'_1, \ldots, B'_{k-1} \vdash B_k \Rightarrow \mathscr{B}$ and $B'_1, \ldots, B'_{k-1} \vdash \neg B_k \Rightarrow \mathscr{B}$. Then by Lemma 1.11(g) and MP, $B'_1, \ldots, B'_{k-1} \vdash \mathscr{B}$. Similarly, B_{k-1} may be chosen to be T or F and, again applying the deduction theorem, Lemma 1.11(g) and MP, we can eliminate B'_{k-1} just as we eliminated B'_k. After k such steps, we finally obtain $\vdash \mathscr{B}$.

COROLLARY 1.15

If $\mathscr{C}$ is an expression involving the signs $\neg$, $\Rightarrow$, $\wedge$, $\vee$ and $\Leftrightarrow$ that is an abbreviation for a wf $\mathscr{B}$ of L, then $\mathscr{C}$ is a tautology if and only if $\mathscr{B}$ is a theorem of L.

Proof

In definitions (D1)–(D3), the abbreviating formulas replace wfs to which they are logically equivalent. Hence, by Proposition 1.4, $\mathscr{B}$ and $\mathscr{C}$ are logically equivalent, and $\mathscr{C}$ is a tautology if and only if $\mathscr{B}$ is a tautology. The corollary now follows from Propositions 1.12 and 1.14.

COROLLARY 1.16

The system L is consistent; that is, there is no wf $\mathscr{B}$ such that both $\mathscr{B}$ and $\neg\mathscr{B}$ are theorems of L.

Proof

By Proposition 1.12, every theorem of L is a tautology. The negation of a tautology cannot be a tautology and, therefore, it is impossible for both $\mathscr{B}$ and $\neg\mathscr{B}$ to be theorems of L.

Notice that L is consistent if and only if not all wfs of L are theorems. In fact, if L is consistent, then there are wfs that are not theorems (e.g., the negations of theorems). On the other hand, by Lemma 1.11(c), $\vdash_L \neg\mathscr{B} \Rightarrow (\mathscr{B} \Rightarrow \mathscr{C})$, and so, if L were inconsistent, that is, if some wf $\mathscr{B}$ and

its negation $\neg\mathscr{B}$ were provable, then by MP any wf $\mathscr{C}$ would be provable. (This equivalence holds for any theory that has *modus ponens* as a rule of inference and in which Lemma 1.11(c) is provable.) A theory in which not all wfs are theorems is said to be *absolutely consistent*, and this definition is applicable even to theories that do not contain a negation sign.

Exercise

1.50 Let $\mathscr{B}$ be a statement form that is not a tautology. Let L^+ be the formal theory obtained from L by adding as new axioms all wfs obtainable from $\mathscr{B}$ by substituting arbitrary statement forms for the statement letters in $\mathscr{B}$, with the same form being substituted for all occurrences of a statement letter. Show that L^+ is inconsistent.

1.5 INDEPENDENCE. MANY-VALUED LOGICS

A subset Y of the set of axioms of a theory is said to be *independent* if some wf in Y cannot be proved by means of the rules of inference from the set of those axioms not in Y.

PROPOSITION 1.17

Each of the axiom schemes (A1)–(A3) is independent.

Proof

To prove the independence of axiom schema (A1), consider the following tables:

A	$\neg A$
0	1
1	1
2	0

A	B	$A \Rightarrow B$
0	0	0
1	0	2
2	0	0
0	1	2
1	1	2
2	1	0
0	2	2
1	2	0
2	2	0

For any assignment of the values 0, 1 and 2 to the statement letters of a wf $\mathscr{B}$, these tables determine a corresponding value of $\mathscr{B}$. If $\mathscr{B}$ always takes the value 0, $\mathscr{B}$ is called *select*. Modus ponens preserves selectness, since it is easy to check that, if $\mathscr{B}$ and $\mathscr{B} \Rightarrow \mathscr{C}$ are select, so is $\mathscr{C}$. One can also verify that all

instances of axiom schemas (A2) and (A3) are select. Hence, any wf derivable from (A2) and (A3) by modus ponens is select. However, $A_1 \Rightarrow (A_2 \Rightarrow A_1)$, which is an instance of (A1), is not select, since it takes the value 2 when A_1 is 1 and A_2 is 2.

To prove the independence of axiom schema (A2), consider the following tables:

A	$\neg A$
0	1
1	0
2	1

A	B	$A \Rightarrow B$
0	0	0
1	0	0
2	0	0
0	1	2
1	1	2
2	1	0
0	2	1
1	2	0
2	2	0

Let us call a wf that always takes the value 0 according to these tables *grotesque*. Modus ponens preserves grotesqueness and it is easy to verify that all instances of (A1) and (A3) are grotesque. However, the instance $(A_1 \Rightarrow (A_2 \Rightarrow A_3)) \Rightarrow ((A_1 \Rightarrow A_2) \Rightarrow (A_1 \Rightarrow A_3))$ of (A2) takes the value 2 when A_1 is 0, A_2 is 0, and A_3 is 1 and, therefore, is not grotesque.

The following argument proves the independence of (A3). Let us call a wf $\mathscr{B}$ *super* if the wf $h(\mathscr{B})$ obtained by erasing all negation signs in $\mathscr{B}$ is a tautology. Each instance of axiom schemas (A1) and (A2) is super. Also, modus ponens preserves the property of being super; for if $h(\mathscr{B} \Rightarrow \mathscr{C})$ and $h(\mathscr{B})$ are tautologies, then $h(\mathscr{C})$ is a tautology. (Just note that $h(\mathscr{B} \Rightarrow \mathscr{C})$ is $h(\mathscr{B}) \Rightarrow h(\mathscr{C})$ and use Proposition 1.2.) Hence, every wf $\mathscr{B}$ derivable from (A1) and (A2) by modus ponens is super. But $h((\neg A_1 \Rightarrow \neg A_1) \Rightarrow ((\neg A_1 \Rightarrow A_1) \Rightarrow A_1))$ is $(A_1 \Rightarrow A_1) \Rightarrow ((A_1 \Rightarrow A_1) \Rightarrow A_1)$, which is not a tautology. Therefore, $(\neg A_1 \Rightarrow \neg A_1) \Rightarrow ((\neg A_1 \Rightarrow A_1) \Rightarrow A_1)$, an instance of (A3), is not super and is thereby not derivable from (A1) and (A2) by modus ponens.

The idea used in the proof of the independence of axiom schemas (A1) and (A2) may be generalized to the notion of a *many-valued logic*. Select a positive integer n, call the numbers $0, 1, \ldots, n$ *truth values*, and choose a number m such that $0 \leqslant m < n$. The numbers $0, 1, \ldots, m$ are called *designated values*. Take a finite number of 'truth tables' representing functions from sets of the form $\{0, 1, \ldots, n\}^k$ into $\{0, 1, \ldots, n\}$. For each truth table, introduces a sign, called the corresponding *connective*. Using these connectives and statement letters, we may construct 'statement forms', and every such statement form containing j distinct letters determines a 'truth function' from $\{0, 1, \ldots, n\}^j$ into $\{0, 1, \ldots, n\}$. A statement form whose corresponding truth function takes only designated values is said to be *exceptional*. The numbers m and n and the basic truth tables are said to

define a (finite) *many-valued logic* M. A formal theory involving statement letters and the connectives of M is said to be *suitable* for M if and only if the theorems of the theory coincide with the exceptional statement forms of M. All these notions obviously can be generalized to the case of an infinite number of truth values. If $n = 1$ and $m = 0$ and the truth tables are those given for $\neg$ and $\Rightarrow$ in section 1.1, then the corresponding two-valued logic is that studied in this chapter. The exceptional wfs in this case were called tautologies. The system L is suitable for this logic, as proved in Propositions 1.12 and 1.14. In the proofs of the independence of axiom schemas (A1) and (A2), two three-valued logics were used.

Exercises

1.51 Prove the independence of axiom schema (A3) by constructing appropriate 'truth tables' for $\neg$ and $\Rightarrow$.

1.52 (McKinsey and Tarski, 1948) Consider the axiomatic theory P in which there is exactly one binary connective $*$, the only rule of inference is modus ponens (that is, $\mathscr{C}$ follows from $\mathscr{B}$ and $\mathscr{B} * \mathscr{C}$), and the axioms are all wfs of the form $\mathscr{B} * \mathscr{B}$. Show that P is not suitable for any (finite) many-valued logic.

1.53 For any (finite) many-valued logic M, prove that there is an axiomatic theory suitable for M.

Further information about many-valued logics can be found in Rosser and Turquette (1952), Rescher (1969), Bolc and Borowik (1992) and Malinowski (1993).

1.6 OTHER AXIOMATIZATIONS

Although the axiom system L is quite simple, there are many other systems that would do as well. We can use, instead of $\neg$ and $\Rightarrow$, any collection of primitive connectives as long as these are adequate for the definition of all other truth-functional connectives.

Examples

L_1: $\vee$ and $\neg$ are the primitive conectives. We use $\mathscr{B} \Rightarrow \mathscr{C}$ as an abbreviation for $\neg\mathscr{B} \vee \mathscr{C}$. We have four axiom schemas: (1) $\mathscr{B} \vee \mathscr{B} \Rightarrow \mathscr{B}$; (2) $\mathscr{B} \Rightarrow \mathscr{B} \vee \mathscr{C}$; (3) $\mathscr{B} \vee \mathscr{C} \Rightarrow \mathscr{C} \vee \mathscr{B}$; and (4) $(\mathscr{C} \Rightarrow \mathscr{D}) \Rightarrow (\mathscr{B} \vee \mathscr{C} \Rightarrow \mathscr{B} \vee \mathscr{D})$. The only rule of inference is modus ponens. Here and below we use the usual rules for eliminating parentheses. This system is developed in Hilbert and Ackermann (1950).

L_2: $\wedge$ and $\neg$ are the primitive connectives. $\mathscr{B} \Rightarrow \mathscr{C}$ is an abbreviation for $\neg(\mathscr{B} \wedge \neg\mathscr{C})$. There are three axiom schemas: (1) $\mathscr{B} \Rightarrow (\mathscr{B} \wedge \mathscr{B})$;

(2) $\mathscr{B} \wedge \mathscr{C} \Rightarrow \mathscr{B}$; and (3) $(\mathscr{B} \Rightarrow \mathscr{C}) \Rightarrow (\neg(\mathscr{C} \wedge \mathscr{D}) \Rightarrow \neg(\mathscr{D} \wedge \mathscr{B}))$. Modus ponens is the only rule of inference. Consult Rosser (1953) for a detailed study.

L_3: This is just like our original system L except that, instead of the axiom schemas (A1)–(A3), we have three specific axioms: (1) $A_1 \Rightarrow (A_2 \Rightarrow A_1)$; (2) $(A_1 \Rightarrow (A_2 \Rightarrow A_3)) \Rightarrow ((A_1 \Rightarrow A_2) \Rightarrow (A_1 \Rightarrow A_2))$; and (3) $(\neg A_2 \Rightarrow \neg A_1) \Rightarrow ((\neg A_2 \Rightarrow A_1) \Rightarrow A_2)$. In addition to modus ponens, we have a substitution rule: we may substitute any wf for all occurrences of a statement letter in a given wf.

L_4: The primitive connectives are $\Rightarrow$, $\wedge$, $\vee$ and $\neg$. Modus ponens is the only rule, and we have ten axiom schemas: (1) $\mathscr{B} \Rightarrow (\mathscr{C} \Rightarrow \mathscr{B})$; (2) $(\mathscr{B} \Rightarrow (\mathscr{C} \Rightarrow \mathscr{D})) \Rightarrow ((\mathscr{B} \Rightarrow \mathscr{C}) \Rightarrow (\mathscr{B} \Rightarrow \mathscr{D}))$; (3) $\mathscr{B} \wedge \mathscr{C} \Rightarrow \mathscr{B}$; (4) $\mathscr{B} \wedge \mathscr{C} \Rightarrow \mathscr{C}$; (5) $\mathscr{B} \Rightarrow (\mathscr{C} \Rightarrow (\mathscr{B} \wedge \mathscr{C}))$; (6) $\mathscr{B} \Rightarrow (\mathscr{B} \vee \mathscr{C})$; (7) $\mathscr{C} \Rightarrow (\mathscr{B} \vee \mathscr{C})$; (8) $(\mathscr{B} \Rightarrow \mathscr{D}) \Rightarrow ((\mathscr{C} \Rightarrow \mathscr{D}) \Rightarrow (\mathscr{B} \vee \mathscr{C} \Rightarrow \mathscr{D}))$; (9) $(\mathscr{B} \Rightarrow \mathscr{C}) \Rightarrow ((\mathscr{B} \Rightarrow \neg\mathscr{C}) \Rightarrow \neg\mathscr{B})$; and (10) $\neg\neg\mathscr{B} \Rightarrow \mathscr{B}$. This system is discussed in Kleene (1952).

Axiomatizations can be found for the propositional calculus that contain only one axiom schema. For example, if $\neg$ and $\Rightarrow$ are the primitive connectives and modus ponens the only rule of inference, then the axiom schema

$$[(((\mathscr{B} \Rightarrow \mathscr{C}) \Rightarrow (\neg\mathscr{D} \Rightarrow \neg\mathscr{E})) \Rightarrow \mathscr{D}) \Rightarrow \mathscr{F}] \Rightarrow [(\mathscr{F} \Rightarrow \mathscr{B}) \Rightarrow (\mathscr{E} \Rightarrow \mathscr{B})]$$

is sufficient (Meredith, 1953). Another single-axiom formulation, due to Nicod (1917), uses only alternative denial $|$. Its rule of inference is: $\mathscr{D}$ follows from $\mathscr{B} \mid (\mathscr{C} \mid \mathscr{D})$ and $\mathscr{B}$, and its axiom schema is

$$(\mathscr{B}|(\mathscr{C}|\mathscr{D})) \mid \{[\mathscr{E}|(\mathscr{E}|\mathscr{E})] \mid [(\mathscr{F}|\mathscr{C})|((\mathscr{B}|\mathscr{F})|(\mathscr{B}|\mathscr{F}))]\}$$

Further information, including historical background, may be found in Church (1956) and in a paper by Lukasiewicz and Tarski in Tarski (1956, IV).

Exercises

1.54 (Hilbert and Ackermann, 1950) Prove the following results about the theory L_1.

(a) $\mathscr{B} \Rightarrow \mathscr{C} \vdash_{L_1} \mathscr{D} \vee \mathscr{B} \Rightarrow \mathscr{D} \vee \mathscr{C}$

(b) $\vdash_{L_1} (\mathscr{B} \Rightarrow \mathscr{C}) \Rightarrow ((\mathscr{D} \Rightarrow \mathscr{B}) \Rightarrow (\mathscr{D} \Rightarrow \mathscr{C}))$

(c) $\mathscr{D} \Rightarrow \mathscr{B}$, $\mathscr{B} \Rightarrow \mathscr{C} \vdash_{L_1} \mathscr{D} \Rightarrow \mathscr{C}$

(d) $\vdash_{L_1} \mathscr{B} \Rightarrow \mathscr{B}$ (i.e., $\vdash_{L_1} \neg\mathscr{B} \vee \mathscr{B}$)

(e) $\vdash_{L_1} \mathscr{B} \vee \neg\mathscr{B}$

(f) $\vdash_{L_1} \mathscr{B} \Rightarrow \neg\neg\mathscr{B}$

(g) $\vdash_{L_1} \neg\mathscr{B} \Rightarrow (\mathscr{B} \Rightarrow \mathscr{C})$

(h) $\vdash_{L_1} \mathscr{B} \vee (\mathscr{C} \vee \mathscr{D}) \Rightarrow ((\mathscr{C} \vee (\mathscr{B} \vee \mathscr{D})) \vee \mathscr{B})$

(i) $\vdash_{L_1} (\mathscr{C} \vee (\mathscr{B} \vee \mathscr{D})) \vee \mathscr{B} \Rightarrow \mathscr{C} \vee (\mathscr{B} \vee \mathscr{D})$

(j) $\vdash_{L_1} \mathscr{B} \vee (\mathscr{C} \vee \mathscr{D}) \Rightarrow \mathscr{C} \vee (\mathscr{B} \vee \mathscr{D})$

(k) $\vdash_{L_1} (\mathscr{B} \Rightarrow (\mathscr{C} \Rightarrow \mathscr{D})) \Rightarrow (\mathscr{C} \Rightarrow (\mathscr{B} \Rightarrow \mathscr{D}))$
(l) $\vdash_{L_1} (\mathscr{D} \Rightarrow \mathscr{B}) \Rightarrow ((\mathscr{B} \Rightarrow \mathscr{C}) \Rightarrow (\mathscr{D} \Rightarrow \mathscr{C}))$
(m) $\mathscr{B} \Rightarrow (\mathscr{C} \Rightarrow \mathscr{D}),\ \mathscr{B} \Rightarrow \mathscr{C} \vdash_{L_1} \mathscr{B} \Rightarrow (\mathscr{B} \Rightarrow \mathscr{D})$
(n) $\mathscr{B} \Rightarrow (\mathscr{C} \Rightarrow \mathscr{D}),\ \mathscr{B} \Rightarrow \mathscr{C} \vdash_{L_1} \mathscr{B} \Rightarrow \mathscr{D}$
(o) If $\Gamma,\ \mathscr{B} \vdash_{L_1} \mathscr{C}$, then $\Gamma \vdash_{L_1} \mathscr{B} \Rightarrow \mathscr{C}$ (deduction theorem)
(p) $\mathscr{C} \Rightarrow \mathscr{B},\ \neg\mathscr{C} \Rightarrow \mathscr{B} \vdash_{L_1} \mathscr{B}$
(q) $\vdash_{L_1} \mathscr{B}$ if and only if $\mathscr{B}$ is a tautology.

1.55 (Rosser, 1953) Prove the following facts about the theory L_2.
(a) $\mathscr{B} \Rightarrow \mathscr{C},\ \mathscr{C} \Rightarrow \mathscr{D} \vdash_{L_2} \neg(\neg\mathscr{D} \wedge \mathscr{B})$
(b) $\vdash_{L_2} \neg(\neg\mathscr{B} \wedge \mathscr{B})$
(c) $\vdash_{L_2} \neg\neg\mathscr{B} \Rightarrow \mathscr{B}$
(d) $\vdash_{L_2} \neg(\mathscr{B} \wedge \mathscr{C}) \Rightarrow (\mathscr{C} \Rightarrow \neg\mathscr{B})$
(e) $\vdash_{L_2} \mathscr{B} \Rightarrow \neg\neg\mathscr{B}$
(f) $\vdash_{L_2} (\mathscr{B} \Rightarrow \mathscr{C}) \Rightarrow (\neg\mathscr{C} \Rightarrow \neg\mathscr{B})$
(g) $\neg\mathscr{B} \Rightarrow \neg\mathscr{C} \vdash_{L_2} \mathscr{C} \Rightarrow \mathscr{B}$
(h) $\mathscr{B} \Rightarrow \mathscr{C} \vdash_{L_2} \mathscr{D} \wedge \mathscr{B} \Rightarrow \mathscr{C} \wedge \mathscr{D}$
(i) $\mathscr{B} \Rightarrow \mathscr{C},\ \mathscr{C} \Rightarrow \mathscr{D},\ \mathscr{D} \Rightarrow \mathscr{E} \vdash_{L_2} \mathscr{B} \Rightarrow \mathscr{E}$
(j) $\vdash_{L_2} \mathscr{B} \Rightarrow \mathscr{B}$
(k) $\vdash_{L_2} \mathscr{B} \wedge \mathscr{C} \Rightarrow \mathscr{C} \wedge \mathscr{B}$
(l) $\mathscr{B} \Rightarrow \mathscr{C},\ \mathscr{C} \Rightarrow \mathscr{D} \vdash_{L_2} \mathscr{B} \Rightarrow \mathscr{D}$
(m) $\mathscr{B} \Rightarrow \mathscr{C},\ \mathscr{D} \Rightarrow \mathscr{E} \vdash_{L_2} \mathscr{B} \wedge \mathscr{D} \Rightarrow \mathscr{C} \wedge \mathscr{E}$
(n) $\mathscr{C} \Rightarrow \mathscr{D} \vdash_{L_2} \mathscr{B} \wedge \mathscr{C} \Rightarrow \mathscr{B} \wedge \mathscr{D}$
(o) $\vdash_{L_2} (\mathscr{B} \Rightarrow (\mathscr{C} \Rightarrow \mathscr{D})) \Rightarrow ((\mathscr{B} \wedge \mathscr{C}) \Rightarrow \mathscr{D})$
(p) $\vdash_{L_2} ((\mathscr{B} \wedge \mathscr{C}) \Rightarrow \mathscr{D}) \Rightarrow (\mathscr{B} \Rightarrow (\mathscr{C} \Rightarrow \mathscr{D}))$
(q) $\mathscr{B} \Rightarrow \mathscr{C},\ \mathscr{B} \Rightarrow (\mathscr{C} \Rightarrow \mathscr{D}) \vdash_{L_2} \mathscr{B} \Rightarrow \mathscr{D}$
(r) $\vdash_{L_2} \mathscr{B} \Rightarrow (\mathscr{C} \Rightarrow \mathscr{B} \wedge \mathscr{C})$
(s) $\vdash_{L_2} \mathscr{B} \Rightarrow (\mathscr{C} \Rightarrow \mathscr{B})$
(t) If $\Gamma,\ \mathscr{B} \vdash_{L_2} \mathscr{C}$, then $\Gamma \vdash_{L_2} \mathscr{B} \Rightarrow \mathscr{C}$ (deduction theorem)
(u) $\vdash_{L_2} (\neg\mathscr{B} \Rightarrow \mathscr{B}) \Rightarrow \mathscr{B}$
(v) $\mathscr{B} \Rightarrow \mathscr{C},\ \neg\mathscr{B} \Rightarrow \mathscr{C} \vdash_{L_2} \mathscr{C}$
(w) $\vdash_{L_2} \mathscr{B}$ if and only if $\mathscr{B}$ is a tautology.

1.56 Show that the theory L_3 has the same theorems as the theory L.

1.57 (Kleene, 1952) Derive the following facts about the theory L_4.
(a) $\vdash_{L_4} \mathscr{B} \Rightarrow \mathscr{B}$
(b) If $\Gamma,\ \mathscr{B} \vdash_{L_4} \mathscr{C}$, then $\Gamma \vdash_{L_4} \mathscr{B} \Rightarrow \mathscr{C}$ (deduction theorem)
(c) $\mathscr{B} \Rightarrow \mathscr{C},\ \mathscr{C} \Rightarrow \mathscr{D} \vdash_{L_4} \mathscr{B} \Rightarrow \mathscr{D}$
(d) $\vdash_{L_4} (\mathscr{B} \Rightarrow \mathscr{C}) \Rightarrow (\neg\mathscr{C} \Rightarrow \neg\mathscr{B})$
(e) $\mathscr{B},\ \neg\mathscr{B} \vdash_{L_4} \mathscr{C}$
(f) $\vdash_{L_4} \mathscr{B} \Rightarrow \neg\neg\mathscr{B}$
(g) $\vdash_{L_4} \neg\mathscr{B} \Rightarrow (\mathscr{B} \Rightarrow \mathscr{C})$
(h) $\vdash_{L_4} \mathscr{B} \Rightarrow (\neg\mathscr{C} \Rightarrow \neg(\mathscr{B} \Rightarrow \mathscr{C}))$
(i) $\vdash_{L_4} \neg\mathscr{B} \Rightarrow (\neg\mathscr{C} \Rightarrow \neg(\mathscr{B} \vee \mathscr{C}))$
(j) $\vdash_{L_4} (\neg\mathscr{C} \Rightarrow \mathscr{B}) \Rightarrow ((\mathscr{C} \Rightarrow \mathscr{B}) \Rightarrow \mathscr{B})$
(k) $\vdash_{L_4} \mathscr{B}$ if and only if $\mathscr{B}$ is a tautology.

1.58[D] Consider the following axiomatization of the propositional calculus $\mathscr{L}$ (due to Lukasiewicz). $\mathscr{L}$ has the same wfs as our system L. Its only rule of inference is modus ponens. Its axiom schemas are:

(a) $(\neg\mathscr{B} \Rightarrow \mathscr{B}) \Rightarrow \mathscr{B}$
(b) $\mathscr{B} \Rightarrow (\neg\mathscr{B} \Rightarrow \mathscr{C})$
(c) $(\mathscr{B} \Rightarrow \mathscr{C}) \Rightarrow ((\mathscr{C} \Rightarrow \mathscr{D}) \Rightarrow (\mathscr{B} \Rightarrow \mathscr{D}))$

Prove that a wf $\mathscr{B}$ of $\mathscr{L}$ is provable in $\mathscr{L}$ if and only if $\mathscr{B}$ is a tautology. [*Hint*: Show that L and $\mathscr{L}$ have the same theorems. However, remember that none of the results proved about L (such as Propositions 1.8–1.13) automatically carries over to $\mathscr{L}$. In particular, the deduction theorem is not available until it is proved for $\mathscr{L}$.]

1.59 Show that axiom schema (A3) of L can be replaced by the schema $(\neg\mathscr{B} \Rightarrow \neg\mathscr{C}) \Rightarrow (\mathscr{C} \Rightarrow \mathscr{B})$ without altering the class of theorems.

1.60 If axiom schema (10) of L_4 is replaced by the schema (10)′: $\neg\mathscr{B} \Rightarrow (\mathscr{B} \Rightarrow \mathscr{C})$, then the new system L_I is called the *intuitionistic* propositional calculus.[†] Prove the following results about L_I.

(a) Consider an $(n+1)$-valued logic with these connectives: $\neg\mathscr{B}$ is 0 when $\mathscr{B}$ is n, and otherwise it is n; $\mathscr{B} \wedge \mathscr{C}$ has the maximum of the values of $\mathscr{B}$ and $\mathscr{C}$, whereas $\mathscr{B} \vee \mathscr{C}$ has the minimum of these values; and $\mathscr{B} \Rightarrow \mathscr{C}$ is 0 if $\mathscr{B}$ has a value not less than that of $\mathscr{C}$, and otherwise it has the same value as $\mathscr{C}$. If we take 0 as the only designated value, all theorems of L_I are exceptional.

(b) $A_1 \vee \neg A_1$ and $\neg\neg A_1 \Rightarrow A_1$ are not theorems of L_I.

(c) For any m, the wf

$$(A_1 \Leftrightarrow A_2) \vee \ldots \vee (A_1 \Leftrightarrow A_m) \vee (A_2 \Leftrightarrow A_3) \vee \ldots \vee (A_2 \Leftrightarrow A_m) \vee \ldots \vee (A_{m-1} \Leftrightarrow A_m)$$

is not a theorem of L_I

(d) (Gödel, 1933) L_I is not suitable for any finite many-valued logic.

(e) (i) If $\Gamma, \mathscr{B} \vdash_{L_I} \mathscr{C}$, then $\Gamma \vdash_{L_I} \mathscr{B} \Rightarrow \mathscr{C}$ (deduction theorem)
(ii) $\mathscr{B} \Rightarrow \mathscr{C}, \mathscr{C} \Rightarrow \mathscr{D} \vdash_{L_I} \mathscr{B} \Rightarrow \mathscr{D}$
(iii) $\vdash_{L_I} \mathscr{B} \Rightarrow \neg\neg\mathscr{B}$
(iv) $\vdash_{L_I} (\mathscr{B} \Rightarrow \mathscr{C}) \Rightarrow (\neg\mathscr{C} \Rightarrow \neg\mathscr{B})$
(v) $\vdash_{L_I} \mathscr{B} \Rightarrow (\neg\mathscr{B} \Rightarrow \mathscr{C})$
(vi) $\vdash_{L_I} \neg\neg(\neg\neg\mathscr{B} \Rightarrow \mathscr{B})$
(vii) $\neg\neg(\mathscr{B} \Rightarrow \mathscr{C}), \neg\neg\mathscr{B} \vdash_{L_I} \neg\neg\mathscr{C}$

[†]The principal origin of intuitionistic logic was L.E.J. Brouwer's belief that classical logic is wrong. According to Brouwer, $\mathscr{B} \vee \mathscr{C}$ is proved only when a proof of $\mathscr{B}$ or a proof of $\mathscr{C}$ has been found. As a consequence, various tautologies, such as $\mathscr{B} \vee \neg\mathscr{B}$, are not generally acceptable. For further information, consult Brouwer (1976), Heyting (1956), Kleene (1952), Troelstra (1969), and Dummett (1977). Jaśkowski (1936) showed that L_I is suitable for a many-valued logic with denumerably many values.

(viii) $\vdash_{L_I} \neg\neg\neg\mathscr{B} \Rightarrow \neg\mathscr{B}$

(ix) $\vdash_{L_I} \neg\neg\mathscr{B}$ if and only if $\mathscr{B}$ is a tautology.

(x) $\vdash_{L_I} \neg\mathscr{B}$ if and only if $\neg\mathscr{B}$ is a tautology.

(xi) If $\mathscr{B}$ has $\wedge$ and $\neg$ as its only connectives, then $\vdash_{L_I} \mathscr{B}$ if and only if $\mathscr{B}$ is a tautology.

1.61 Let $\mathscr{B}$ and $\mathscr{C}$ be in the relation R if and only if $\vdash_L \mathscr{B} \Leftrightarrow \mathscr{C}$. Show that R is an equivalence relation. Given equivalence classes $[\mathscr{B}]$ and $[\mathscr{C}]$, let $[\mathscr{B}] \cup [\mathscr{C}] = [\mathscr{B} \vee \mathscr{C}]$, $[\mathscr{B}] \cap [\mathscr{C}] = [\mathscr{B} \wedge \mathscr{C}]$, and $\overline{[\mathscr{B}]} = [\neg\mathscr{B}]$. Show that the equivalence classes under R form a Boolean algebra with respect to $\cap$, $\cup$ and $^-$, called the *Lindenbaum algebra* L^* determined by L. The element 0 of L^* is the equivalence class consisting of all contradictions (i.e., negations of tautologies). The unit element 1 of L^* is the equivalence class consisting of all tautologies. Notice that $\vdash_L \mathscr{B} \Rightarrow \mathscr{C}$ if and only if $[\mathscr{B}] \leqslant [\mathscr{C}]$ in L^*, and that $\vdash_L \mathscr{B} \Leftrightarrow \mathscr{C}$ if and only if $[\mathscr{B}] = [\mathscr{C}]$. Show that a Boolean function f (built up from variables, 0, and 1, using $\cup$, $\cap$ and $^-$) is equal to the constant function 1 in all Boolean algebras if and only if $\vdash_L f^{\#}$, where $f^{\#}$ is obtained from f by changing $\cup$, $\cap$, $^-$, 0 and 1 to $\vee$, $\wedge$, $\neg$, $A_1 \wedge \neg A_1$, and $A_1 \vee \neg A_1$, respectively.

2 Quantification Theory

2.1 QUANTIFIERS

There are various kinds of logical inference that cannot be justified on the basis of the propositional calculus; for example:

1. Any friend of Martin is a friend of John.
   Peter is not John's friend.
   Hence, Peter is not Martin's friend.
2. All human beings are rational.
   Some animals are human beings.
   Hence, some animals are rational.
3. The successor of an even integer is odd.
   2 is an even integer.
   Hence, the successor of 2 is odd.

The correctness of these inferences rests not only upon the meanings of the truth-functional connectives, but also upon the meaning of such expressions as 'any', 'all' and 'some', and other linguistic constructions.

In order to make the structure of complex sentences more transparent, it is convenient to introduce special notation to represent frequently occurring expressions. If $P(x)$ asserts that x has the property P, then $(\forall x)P(x)$ means that property P holds for all x or, in other words, that everything has the property P. On the other hand, $(\exists x)P(x)$ means that some x has the property P – that is, that there is at least one object having the property P. In $(\forall x)P(x)$, '$(\forall x)$' is called a *universal quantifier*; in $(\exists x)P(x)$, '$(\exists x)$' is called an *existential quantifier*. The study of quantifiers and related concepts is the principal subject of this chapter.

Examples

1′. Inference 1 above can be represented symbolically:

$$\frac{\begin{array}{l}(\forall x)(F(x,m) \Rightarrow F(x,j)) \\ \neg F(p,j)\end{array}}{\neg F(p,m)}$$

Here, $F(x,y)$ means that x is a friend of y, while m, j and p denote Martin, John, and Peter, respectively.

2′. Inference 2 becomes:

$$\frac{\begin{array}{l}(\forall x)(H(x) \Rightarrow R(x)) \\ (\exists x)(A(x) \wedge H(x))\end{array}}{(\exists x)(A(x) \wedge R(x))}$$

Here, H, R and A designate the properties of being human, rational, and an animal, respectively.

3′. Inference 3 can be symbolized as follows:

$$\frac{\begin{array}{l}(\forall x)(I(x) \wedge E(x) \Rightarrow D(s(x))) \\ I(b) \wedge E(b)\end{array}}{D(s(b))}$$

Here, I, E and D designate respectively the properties of being an integer, even and odd; $s(x)$ denotes the successor of x; and b denotes the integer 2.

Notice that the validity of these inferences does not depend upon the particular meanings of $F, m, j, p, H, R, A, I, E, D, s$ and b.

Just as statement forms were used to indicate logical structure dependent upon the logical connectives, so also the form of inferences involving quantifiers, such as inferences 1–3, can be represented abstractly, as in 1′ – 3′. For this purpose, we shall use commas, parentheses, the symbols $\neg$ and $\Rightarrow$ of the propositional calculus, the universal quantifier symbol $\forall$, and the following groups of symbols:

Individual variables: $x_1, x_2, \ldots, x_n, \ldots$
Individual constants: $a_1, a_2, \ldots, a_n, \ldots$
Predicate letters: A_k^n (n and k are any positive integers)
Function letters: f_k^n (n and k are any positive integers)

The positive integer n that is a superscript of a predicate letter A_k^n or of a function letter f_k^n indicates the number of arguments, whereas the subscript k is just an indexing number to distinguish different predicate or function letters with the same number of arguments.†

In the preceding examples, x plays the role of an individual variable; m, j, p and b play the role of individual constants; F is a binary predicate letter (i.e., a predicate letter with two arguments); H, R, A, I, E and D are monadic predicate letters (i.e., predicate letters with one argument); and s is a function letter with one argument.

The function letters applied to the variables and individual constants generate the *terms*:

†For example, in arithmetic both addition and multiplication take two arguments. So, we would use one function letter, say f_1^2, for addition, and a different function letter, say f_2^2, for multiplication.

1. Variables and individual constants are terms.
2. If f_k^n is a function letter and $t_1, t_2, \ldots, t_n$ are terms, then $f_k^n(t_1, t_2, \ldots, t_n)$ is a term.
3. An expression is a term only if it can be shown to be a term on the basis of conditions 1 and 2.

Terms correspond to what in ordinary languages are nouns and noun phrases – for example, 'two', 'two plus three', and 'two plus x'.

The predicate letters applied to terms yield the *atomic formulas*; that is, if A_k^n is a predicate letter and $t_1, t_2, \ldots, t_n$ are terms, then $A_k^n(t_1, t_2, \ldots, t_n)$ is an atomic formula.

The *well-formed formulas* (wfs) of quantification theory are defined as follows:

1. Every atomic formula is a wf.
2. If $\mathscr{B}$ and $\mathscr{C}$ are wfs and y is a variable, then $(\neg\mathscr{B})$, $(\mathscr{B} \Rightarrow \mathscr{C})$, and $((\forall y)\mathscr{B})$ are wfs.
3. An expression is a wf only if it can be shown to be a wf on the basis of conditions 1 and 2.

In $((\forall y)\mathscr{B})$, '$\mathscr{B}$' is called the *scope* of the quantifier '$(\forall y)$'. Notice that $\mathscr{B}$ need not contain the variable y. In that case, we understand $((\forall y)\mathscr{B})$ to mean the same thing as $\mathscr{B}$.

The expressions $(\mathscr{B} \wedge \mathscr{C})$, $(\mathscr{B} \vee \mathscr{C})$, and $(\mathscr{B} \Leftrightarrow \mathscr{C})$ are defined as in system L (see page 36). It was unnecessary for us to use the symbol $\exists$ as a primitive symbol because we can define existential quantification as follows:

$((\exists x)\mathscr{B})$ stands for $(\neg((\forall x)(\neg\mathscr{B})))$

This definition is faithful to the meaning of the quantifiers: $\mathscr{B}(x)$ is true for some x if and only if it is not the case that $\mathscr{B}(x)$ is false for all x.†

Parentheses

The same conventions as made in Chapter 1 (page 20) about the omission of parentheses are made here, with the additional convention that quantifiers $(\forall y)$ and $(\exists y)$ rank in strength between $\neg, \wedge, \vee$ and $\Rightarrow, \Leftrightarrow$.

Examples

Parentheses are restored in the following steps.

1. $(\forall x_1)A_1^1(x_1) \Rightarrow A_1^2(x_2, x_1)$
   $((\forall x_1)A_1^1(x_1)) \Rightarrow A_1^2(x_2, x_1)$
   $(((\forall x_1)A_1^1(x_1)) \Rightarrow A_1^2(x_2, x_1))$

†We could have taken $\exists$ as primitive and then defined $((\forall x)\mathscr{B})$ as an abbreviation for $(\neg((\exists x)(\neg\mathscr{B})))$, since $\mathscr{B}(x)$ is true for all x if and only if it is not the case that $\mathscr{B}(x)$ is false for some x.

2. $(\forall x_1)A_1^1(x_1) \vee A_1^2(x_2, x_1)$
   $(\forall x_1)(A_1^1(x_1) \vee A_1^2(x_2, x_1))$
   $((\forall x_1)(A_1^1(x_1)) \vee A_1^2(x_2, x_1)))$
3. $(\forall x_1)(\exists x_2)A_1^2(x_1, x_2)$
   $(\forall x_1)((\exists x_2)A_1^2(x_1, x_2))$
   $((\forall x_1)((\exists x_2)A_1^2(x_1, x_2)))$

Exercises

2.1 Restore parentheses to the following.

(a) $(\forall x_1)A_1^1(x_1) \wedge \neg A_1^1(x_2)$
(b) $(\forall x_2)A_1^1(x_2) \Leftrightarrow A_1^1(x_2)$
(c) $(\forall x_2)(\exists x_1)A_1^2(x_1, x_2)$
(d) $(\forall x_1)(\forall x_3)(\forall x_4)A_1^1(x_1) \Rightarrow A_1^1(x_2) \wedge \neg A_1^1(x_1)$
(e) $(\exists x_1)(\forall x_2)(\exists x_3)A_1^1(x_1) \vee (\exists x_2)\neg(\forall x_3)A_1^2(x_3, x_2)$
(f) $(\forall x_2)\neg A_1^1(x_1) \Rightarrow A_1^3(x_1, x_1, x_2) \vee (\forall x_1)A_1^1(x_1)$
(g) $\neg(\forall x_1)A_1^1(x_1) \Rightarrow (\exists x_2)A_1^1(x_2) \Rightarrow A_1^2(x_1, x_2) \wedge A_1^1(x_2)$

2.2 Eliminate parentheses from the following wfs as far as is possible.

(a) $(((\forall x_1)(A_1^1(x_1) \Rightarrow A_1^1(x_1))) \vee ((\exists x_1)A_1^1(x_1)))$
(b) $((\neg((\exists x_2)(A_1^1(x_2) \vee A_1^1(a_1)))) \Leftrightarrow A_1^1(x_2))$
(c) $(((\forall x_1)(\neg(\neg A_1^1(a_3)))) \Rightarrow (A_1^1(x_1) \Rightarrow A_1^1(x_2)))$

An occurrence of a variable x is said to be *bound* in a wf $\mathscr{B}$ if either it is the occurrence of x in a quantifier '$(\forall x)$' in $\mathscr{B}$ or it lies within the scope of a quantifier '$(\forall x)$' in $\mathscr{B}$. Otherwise, the occurrence is said to be *free* in $\mathscr{B}$.

Examples

1. $A_1^2(x_1, x_2)$
2. $A_1^2(x_1, x_2) \Rightarrow (\forall x_1)A_1^1(x_1)$
3. $(\forall x_1)(A_1^2(x_1, x_2) \Rightarrow (\forall x_1)A_1^1(x_1))$
4. $(\exists x_1)A_1^2(x_1, x_2)$

In Example 1, the single occurrence of x_1 is free. In Example 2, the first occurrence of x_1 is free, but the second and third occurrences are bound. In Example 3, all occurrences of x_1 are bound, and in Example 4 both occurrences of x_1 are bound. (Remember that $(\exists x_1)A_1^2(x_1, x_2)$ is an abbreviation of $\neg(\forall x_1)\neg A_1^2(x_1, x_2)$.) In all four wfs, every occurrence of x_2 is free. Notice that, as in Example 2, a variable may have both free and bound occurrences in the same wf. Also observe that an occurrence of a variable may be bound in some wf $\mathscr{B}$ but free in a subformula of $\mathscr{B}$. For example, the first occurrence of x_1 is free in the wf of Example 2 but bound in the larger wf of Example 3.

A variable is said to be *free* (*bound*) in a wf $\mathscr{B}$ if it has a free (bound) occurrence in $\mathscr{B}$. Thus, a variable may be both free and bound in the same wf; for example, x_1 is free and bound in the wf of Example 2.

Exercises

2.3 Pick out the free and bound occurrences of variables in the following wfs.

(a) $(\forall x_3)(((\forall x_1)A_1^2(x_1, x_2)) \Rightarrow A_1^2(x_3, a_1))$

(b) $(\forall x_2)A_1^2(x_3, x_2) \Rightarrow (\forall x_3)A_1^2(x_3, x_2)$

(c) $((\forall x_2)(\exists x_1)A_1^3(x_1, x_2, f_1^2(x_1, x_2))) \vee \neg(\forall x_1)A_1^2(x_2, f_1^1(x_1))$

2.4 Indicate the free and bound occurrences of all variables in the wfs of Exercises 2.1 and 2.2.

2.5 Indicate the free and bound variables in the wfs of Exercises 2.1–2.3.

We shall often indicate that some of the variables $x_{i_1}, \ldots, x_{i_k}$ are free variables in a wf $\mathscr{B}$ by writing $\mathscr{B}$ as $\mathscr{B}(x_{i_1}, \ldots, x_{i_k})$. This does not mean that $\mathscr{B}$ contains these variables as free variables, nor does it mean that $\mathscr{B}$ does not contain other free variables. This notation is convenient because we can then agree to write as $\mathscr{B}(t_1, \ldots, t_k)$ the result of substituting in $\mathscr{B}$ the terms $t_1, \ldots, t_k$ for all free occurrences (if any) of $x_{i_1}, \ldots, x_{i_k}$, respectively.

If $\mathscr{B}$ is a wf and t is a term, then t is said to be *free for* x_i *in* $\mathscr{B}$ if no free occurrence of x_i in $\mathscr{B}$ lies within the scope of any quantifier $(\forall x_j)$, where x_j is a variable in t. This concept of t being free for x_i in a wf $\mathscr{B}(x_i)$ will have certain technical applications later on. It means that, if t is substituted for all free occurrences (if any) of x_i in $\mathscr{B}(x_i)$, no occurrence of a variable in t becomes a bound occurrence in $\mathscr{B}(t)$.

Examples

1. The term x_2 is free for x_1 in $A_1^1(x_1)$, but x_2 is not free for x_1 in $(\forall x_2)A_1^1(x_1)$.
2. The term $f_1^2(x_1, x_3)$ is free for x_1 in $(\forall x_2)A_1^2(x_1, x_2) \Rightarrow A_1^1(x_1)$ but is not free for x_1 in $(\exists x_3)(\forall x_2)A_1^2(x_1, x_2) \Rightarrow A_1^1(x_1)$.

The following facts are obvious.

1. A term that contains no variables is free for any variable in any wf.
2. A term t is free for any variable in $\mathscr{B}$ if none of the variables of t is bound in $\mathscr{B}$.
3. x_i is free for x_i in any wf.
4. Any term is free for x_i in $\mathscr{B}$ if $\mathscr{B}$ contains no free occurrences of x_i.

Exercises

2.6 Is the term $f_1^2(x_1, x_2)$ free for x_1 in the following wfs?

(a) $A_1^2(x_1, x_2) \Rightarrow (\forall x_2)A_1^1(x_2)$

(b) $((\forall x_2)A_1^2(x_2, a_1)) \vee (\exists x_2)A_1^2(x_1, x_2)$

(c) $(\forall x_1)A_1^2(x_1, x_2)$

(d) $(\forall x_2)A_1^2(x_1, x_2)$

(e) $(\forall x_2)A_1^1(x_2) \Rightarrow A_1^2(x_1, x_2)$

2.7 Justify facts 1–4 above.

When English sentences are translated into formulas, certain general guidelines will be useful:

1. A sentence of the form 'All As are Bs' becomes $(\forall x)(A(x) \Rightarrow B(x))$. For example, *Every mathematician loves music* is translated as $(\forall x)(M(x) \Rightarrow L(x))$, where $M(x)$ means x *is a mathematician* and $L(x)$ means x *loves music*.
2. A sentence of the form 'Some As are Bs' becomes $(\exists x)(A(x) \wedge B(x))$. For example, *Some New Yorkers are friendly* becomes $(\exists x)(N(x) \wedge F(x))$, where $N(x)$ means x *is a New Yorker* and $F(x)$ means x *is friendly*.
3. A sentence of the form 'No As are Bs' becomes $(\forall x)(A(x) \Rightarrow \neg B(x))$.† For example, *No philosopher understands politics* becomes $(\forall x)(P(x) \Rightarrow \neg U(x))$, where $P(x)$ means x *is a philosopher* and $U(x)$ means x *understands politics*.

Let us consider a more complicated example: *Some people respect everyone*. This can be translated as $(\exists x)(P(x) \wedge (\forall y)(P(y) \Rightarrow R(x, y)))$, where $P(x)$ means x *is a person* and $R(x, y)$ means x *respects* y.

Notice that, in informal discussions, to make formulas easier to read we may use lower-case letters u, v, x, y, z instead of our official notation x_i for individual variables, capital letters $A, B, C, \ldots$ instead of our official notation A_k^n for predicate letters, lower-case letters $f, g, h, \ldots$ instead of our official notation f_k^n for function letters, and lower-case letters $a, b, c, \ldots$ instead of our official notation a_i for individual constants.

Exercises

2.8 Translate the following sentences into wfs.

(a) Anyone who is persistent can learn logic.
(b) No politician is honest.
(c) Not all birds can fly.
(d) All birds cannot fly.
(e) x is transcendental only if it is irrational.
(f) Seniors date only juniors.
(g) If anyone can solve the problem, Hilary can.
(h) Nobody loves a loser.
(i) Nobody in the statistics class is smarter than everyone in the logic class.
(j) John hates all people who do not hate themselves.
(k) Everyone loves somebody and no one loves everybody, or somebody loves everybody and someone loves nobody.
(l) You can fool some of the people all of the time, and you can fool all the people some of the time, but you can't fool all the people all the time.

†As we shall see later, this is equivalent to $\neg(\exists x)(A(x) \wedge B(x))$.

(m) Any sets that have the same members are equal.
(n) Anyone who knows Julia loves her.
(o) There is no set belonging to precisely those sets that do not belong to themselves.
(p) There is no barber who shaves precisely those men who do not shave themselves.

2.9 Translate the following into everyday English. Note that everyday English does not use variables.

(a) $(\forall x)(M(x) \wedge (\forall y)\neg W(x,y) \Rightarrow U(x))$, where $M(x)$ means x *is a man*, $W(x,y)$ means x *is married to* y, and $U(x)$ means x *is unhappy*.
(b) $(\forall x)(V(x) \wedge P(x) \Rightarrow A(x,b))$, where $V(x)$ means x *is an even integer*, $P(x)$ means x *is a prime integer*, $A(x,y)$ means $x = y$, and b denotes 2.
(c) $\neg(\exists y)(I(y) \wedge (\forall x)(I(x) \Rightarrow L(x,y)))$, where $I(y)$ means y *is an integer* and $L(x,y)$ means $x \leqslant y$.
(d) In the following wfs, $A_1^1(x)$ means x *is a person* and $A_1^2(x,y)$ means x *hates* y.
  (i) $(\exists x)(A_1^1(x) \wedge (\forall y)(A_1^1(y) \Rightarrow A_1^2(x,y)))$
  (ii) $(\forall x)(A_1^1(x) \Rightarrow (\forall y)(A_1^1(y) \Rightarrow A_1^2(x,y)))$
  (iii) $(\exists x)(A_1^1(x) \wedge (\forall y)(A_1^1(y) \Rightarrow (A_1^2(x,y) \Leftrightarrow A_1^2(y,y))))$
(e) $(\forall x)(H(x) \Rightarrow (\exists y)(\exists z)(\neg A(y,z) \wedge (\forall u)(P(u,x) \Leftrightarrow (A(u,y) \vee A(u,z)))))$, where $H(x)$ means x *is a person*, $A(u,v)$ means '$u = v$', and $P(u,x)$ means u *is a parent of* x.

2.2 FIRST-ORDER LANGUAGES AND THEIR INTERPRETATIONS. SATISFIABILITY AND TRUTH. MODELS

Well-formed formulas have meaning only when an interpretation is given for the symbols. We usually are interested in interpreting wfs whose symbols come from a specific language. For that reason, we shall define the notion of a *first-order language*.†

†The adjective 'first-order' is used to distinguish the languages we shall study here from those in which there are predicates having other predicates or functions as arguments or in which predicate quantifiers or function quantifiers are permitted, or both. Most mathematical theories can be formalized within first-order languages, although there may be a loss of some of the intuitive content of those theories. Second-order languages are discussed in the appendix on second-order logic. Examples of higher-order languages are studied also in Gödel (1931), Tarski (1933), Church (1940), Hasenjaeger and Scholz (1961) and Van Bentham and Doets (1983). Differences between first-order and higher-order theories are examined in Corcoran (1980).

DEFINITION

A first-order language $\mathscr{L}$ contains the following symbols.

(a) The propositional connectives $\neg$ and $\Rightarrow$, and the universal quantifier symbol $\forall$.
(b) Punctuation marks: the left parenthesis (, the right parenthesis), and the comma.†
(c) Denumerably many individual variables $x_1, x_2, \ldots$.
(d) A finite or denumerable, possibly empty, set of function letters.
(e) A finite or denumerable, possibly empty, set of individual constants.
(f) A non-empty set of predicate letters.

By a *term of* $\mathscr{L}$ we mean a term whose symbols are symbols of $\mathscr{L}$.
By a *wf of* $\mathscr{L}$ we mean a wf whose symbols are symbols of $\mathscr{L}$.

Thus, in a language $\mathscr{L}$, some or all of the function letters and individual constants may be absent, and some (but not all) of the predicate letters may be absent.‡ The individual constants, function letters and predicate letters of a language $\mathscr{L}$ are called the *non-logical constants* of $\mathscr{L}$. Languages are designed in accordance with the subject matter we wish to study. A language for arithmetic might contain function letters for addition and multiplication and a predicate letter for equality, whereas a language for geometry is likely to have predicate letters for equality and the notions of *point* and *line* but no function letters at all.

DEFINITION

Let $\mathscr{L}$ be a first-order language. An *interpretation* M of $\mathscr{L}$ consists of the following ingredients.

(a) A non-empty set D, called the *domain* of the interpretation.
(b) For each predicate letter A_j^n of $\mathscr{L}$, an assignment of an n-place relation $(A_j^n)^{\mathrm{M}}$ in D.
(c) For each function letter f_j^n of $\mathscr{L}$, an assignment of an n-place operation $(f_j^n)^{\mathrm{M}}$ in D (that is, a function from D^n into D).
(d) For each individual constant a_i of $\mathscr{L}$, an assignment of some fixed element $(a_i)^{\mathrm{M}}$ of D.

Given such an interpretation, variables are thought of as ranging over the set D, and $\neg, \Rightarrow$ and quantifiers are given their usual meaning. Remember that an n-place relation in D can be thought of as a subset of D^n, the set of all

†The punctuation marks are not strictly necessary; they can be avoided by redefining the notions of term and wf. However, their use makes it easier to read and comprehend formulas.

‡If there were no predicate letters, there would be no wfs.

n-tuples of elements of D. For example, if D is the set of human beings, then the relation 'father of' can be identified with the set of all ordered pairs $\langle x, y\rangle$ such that x is the father of y.

For a given interpretation of a language $\mathscr{L}$, a wf of $\mathscr{L}$ without free variables (called a *closed wf* or a *sentence*) represents a propostion that is true or false, whereas a wf with free variables may be satisfied (i.e., true) for some values in the domain and not satisfied (i.e., false) for the others.

Examples
Consider the following wfs:

1. $A_1^2(x_1, x_2)$
2. $(\forall x_2)A_1^2(x_1, x_2)$
3. $(\exists x_1)(\forall x_2)A_1^2(x_1, x_2)$

Let us take as domain the set of all positive integers and interpret $A_1^2(y, z)$ as $y \leqslant z$. Then wf 1 represents the expression '$x_1 \leqslant x_2$', which is satisfied by all the ordered pairs $\langle a, b\rangle$ of positive integers such that $a \leqslant b$. Wf 2 represents the expression 'For all positive integers x_2, $x_1 \leqslant x_2$,'† which is satisfied only by the integer 1. Wf 3 is a true sentence asserting that there is a smallest positive integer. If we were to take as domain the set of all integers, then wf 3 would be false.

Exercises

2.10 For the following wfs and for the given interpretations, indicate for what values the wfs are satisfied (if they contain free variables) or whether they are true or false (if they are closed wfs).

(i) $A_1^2(f_1^2(x_1, x_2), a_1)$
(ii) $A_1^2(x_1, x_2) \Rightarrow A_1^2(x_2, x_1)$
(iii) $(\forall x_1)(\forall x_2)(\forall x_3)(A_1^2(x_1, x_2) \wedge A_1^2(x_2, x_3) \Rightarrow A_1^2(x_1, x_3))$

(a) The domain is the set of positive integers, $A_1^2(y, z)$ is $y \geqslant z$, $f_1^2(y, z)$ is $y \cdot z$, and a_1 is 2.
(b) The domain is the set of integers, $A_1^2(y, z)$ is $y = z$, $f_1^2(y, z)$ is $y + z$, and a_1 is 0.
(c) The domain is the set of all sets of integers, $A_1^2(y, z)$ if $y \subseteq z$, $f_1^2(y, z)$ is $y \cap z$, and a_1 is the empty set $\emptyset$.

2.11 Describe in everyday English the assertions determined by the following wfs and interpretations.

(a) $(\forall x)(\forall y)(A_1^2(x, y) \Rightarrow (\exists z)(A_1^1(z) \wedge A_1^2(x, z) \wedge A_1^2(z, y)))$, where the domain D is the set of real numbers, $A_1^2(x, y)$ means $x < y$, and $A_1^1(z)$ means z *is a rational number.*

†In ordinary English, one would say 'x_1 is less than or equal to all positive integers'.

(b) $(\forall x)(A_1^1(x) \Rightarrow (\exists y)(A_2^1(y) \wedge A_1^2(y,x)))$, where D is the set of all days and people, $A_1^1(x)$ means x *is a day*, $A_2^1(y)$ means y *is a sucker*, and $A_1^2(y,x)$ means y *is born on day* x.

(c) $(\forall x)(\forall y)(A_1^1(x) \wedge A_1^1(y) \Rightarrow A_2^1(f_1^2(x,y)))$, where D is the set of integers, $A_1^1(x)$ means x *is odd*, $A_2^1(x)$ means x *is even*, and $f_1^2(x,y)$ denotes $x+y$.

(d) For the following wfs, D is the set of all people and $A_1^2(u,v)$ means u *loves* v.

(i) $(\exists x)(\forall y)(A_1^2(x,y)$

(ii) $(\forall y)(\exists x)A_1^2(x,y)$

(iii) $(\exists x)(\forall y)((\forall z)(A_1^2(y,z) \Rightarrow A_1^2(x,y))$

(iv) $(\exists x)(\forall y)\neg A_1^2(x,y)$

The concepts of satisfiability and truth are intuitively clear, but, following Tarski (1936), we also can provide a rigorous definition. Such a definition is necessary for carrying out precise proofs of many metamathematical results.

Satisfiability will be the fundamental notion, on the basis of which the notion of truth will be defined. Moreover, instead of talking about the n-tuples of objects that satisfy a wf that has n free variables, it is much more convenient from a technical standpoint to deal uniformly with denumerable sequences. What we have in mind is that a denumerable sequence $s = (s_1, s_2, s_3, \ldots)$ is to be thought of as satisfying a wf $\mathscr{B}$ that has $x_{j_1}, x_{j_2} \ldots, x_{j_n}$ as free variables (where $j_1 < j_2 < \ldots < j_n$) if the n-tuple $\langle s_{j_1}, s_{j_2}, \ldots, s_{j_n}\rangle$ satisfies $\mathscr{B}$ in the usual sense. For example, a denumerable sequence $(s_1, s_2, s_3, \ldots)$ of objects in the domain of an interpretation M will turn out to satisfy the wf $A_1^2(x_2, x_5)$ if and only if the ordered pair, $\langle s_2, s_5\rangle$ is in the relation $(A_1^2)^{\mathrm{M}}$ assigned to the predicate letter A_1^2 by the interpretation M.

Let M be an interpretation of a language $\mathscr{L}$ and let D be the domain of M. Let Σ be the set of all denumerable sequences of elements of D. For a wf $\mathscr{B}$ of $\mathscr{L}$, we shall define what it means for a sequence $s = (s_1, s_2, \ldots)$ in Σ to satisfy $\mathscr{B}$ in M. As a preliminary step, for a given s in Σ we shall define a function s^* that assigns to each term t of $\mathscr{L}$ an element $s^*(t)$ in D.

1. If t is a variable x_j, let $s^*(t)$ be s_j.
2. If t is an individual constant a_j, then $s^*(t)$ is the interpretation $(a_j)^{\mathrm{M}}$ of this constant.
3. If f_k^n is a function letter, $(f_k^n)^{\mathrm{M}}$ is the corresponding operation in D, and $t_1, \ldots, t_n$ are terms, then

$$s^*(f_k^n(t_1, \ldots, t_n)) = (f_k^n)^{\mathrm{M}}(s^*(t_1), \ldots, s^*(t_n))$$

Intuitively, $s^*(t)$ is the element of D obtained by substituting, for each j, a name of s_j for all occurrences of x_j in t and then performing the operations of the interpretation corresponding to the function letters of t. For instance, if t is $f_2^2(x_3, f_1^2(x_1, a_1))$ and if the interpretation has the set of integers as its domain, f_2^2 and f_1^2 are interpreted as ordinary multiplication and addition, respectively, and a_1 is interpreted as 2, then, for any sequence $s = (s_1, s_2, \ldots)$

of integers, $s^*(t)$ is the integer $s_3 \cdot (s_1 + 2)$. This is really nothing more than the ordinary way of reading mathematical expressions.

Now we proceed to the definition of satisfaction, which will be an inductive definition.

1. If $\mathscr{B}$ is an atomic wf $A_k^n(t_1, \ldots, t_n)$ and $(A_k^n)^{\mathrm{M}}$ is the corresponding n-place relation of the interpretation, then a sequence $s = (s_1, s_2, \ldots)$ satisfies $\mathscr{B}$ if and only if $(A_k^n)^{\mathrm{M}}(s^*(t_1), \ldots, s^*(t_n))$ – that is, if the n-tuple $\langle s^*(t_1), \ldots, s^*(t_n)\rangle$ is in the relation $(A_k^n)^{\mathrm{M}}$.†
2. s satisfies $\neg\, \mathscr{B}$ if and only if s does not satisfy $\mathscr{B}$.
3. s satisfies $\mathscr{B} \Rightarrow \mathscr{C}$ if and only if s does not satisfy $\mathscr{B}$ or s satisfies $\mathscr{C}$.
4. s satisfies $(\forall x_i)\mathscr{B}$ if and only if every sequence that differs from s in at most the ith component satisfies $\mathscr{B}$.‡

Intuitively, a sequence $s = (s_1, s_2, \ldots)$ satisfies a wf $\mathscr{B}$ if and only if, when, for each i, we replace all free occurrences of x_i (if any) in $\mathscr{B}$ by a symbol representing s_i, the resulting proposition is true under the given interpretation.

Now we can define the notions of truth and falsity of wfs for a given interpretation.

DEFINITIONS

1. A wf $\mathscr{B}$ is *true for the interpretation* M (written $\models_{\mathrm{M}} \mathscr{B}$) if and only if every sequence in Σ satisfies $\mathscr{B}$.
2. $\mathscr{B}$ is said to be *false for* M if and only if no sequence in Σ satisfies $\mathscr{B}$.
3. An interpretation M is said to be a *model* for a set Γ of wfs if and only if every wf in Γ is true for M.

The plausibility of our definition of truth will be strengthened by the fact that we can derive all of the following expected properties I–XI of the notions of truth, falsity and satisfaction. Proofs that are not explicitly given are left to the reader (or may be found in the answer to Exercise 2.12). Most

†For example, if the domain of the interpretation is the set of real numbers, the interpretation of A_1^2 is the relation $\leqslant$, and the interpretation of f_1^1 is the function e^x, then a sequence $s = (s_1, s_2, \ldots)$ of real numbers satisfies $A_1^2(f_1^1(x_2), x_5)$ if and only if $e^{s_2} \leqslant s_5$. If the domain is the set of integers, the interpretation of $A_1^4(x, y, u, v)$ is $x \cdot v = u \cdot y$, and the interpretation of a_1 is 3, then a sequence $s = (s_1, s_2, \ldots)$ of integers satisfies $A_1^4(x_3, a_1, x_1, x_3)$ if and only if $(s_3)^2 = 3s_1$.

‡In other words, a sequence $s = (s_1, s_2, \ldots, s_i, \ldots)$ satisfies $(\forall x_i)\mathscr{B}$ if and only if, for every element c of the domain, the sequence $(s_1, s_2, \ldots, c, \ldots)$ satisfies $\mathscr{B}$. Here, $(s_1, s_2, \ldots, c, \ldots)$ denotes the sequence obtained from $(s_1, s_2, \ldots, s_i, \ldots)$ by replacing the ith component s_i by c. Note also that, if s satisfies $(\forall x_i)\mathscr{B}$, then, as a special case, s satisfies $\mathscr{B}$.

of the results are also obvious if one wishes to use only the ordinary intuitive understanding of the notions of truth, falsity and satisfaction.

(I) (a) $\mathscr{B}$ is false for an interpretation M if and only if $\neg\mathscr{B}$ is true for M.
(b) $\mathscr{B}$ is true for M if and only if $\neg\mathscr{B}$ is false for M.

(II) It is not the case that both $\models_M \mathscr{B}$ and $\models_M \neg\mathscr{B}$; that is, no wf can be both true and false for M.

(III) If $\models_M \mathscr{B}$ and $\models_M \mathscr{B} \Rightarrow \mathscr{C}$, then $\models_M \mathscr{B}$.

(IV) $\mathscr{B} \Rightarrow \mathscr{C}$ is false for M if and only if $\models_M \mathscr{B}$ and $\models_M \neg\mathscr{C}$.

(V)† Consider an interpretation M with domain D.
(a) A sequence s satisfies $\mathscr{B} \wedge \mathscr{C}$ if and only if s satisfies $\mathscr{B}$ and s satisfies $\mathscr{C}$.
(b) s satisfies $\mathscr{B} \vee \mathscr{C}$ if and only if s satisfies $\mathscr{B}$ or s satifies $\mathscr{C}$.
(c) s satisfies $\mathscr{B} \Leftrightarrow \mathscr{C}$ if and only if s satisfies both $\mathscr{B}$ and $\mathscr{C}$ or s satisfies neither $\mathscr{B}$ nor $\mathscr{C}$.
(d) s satisfies $(\exists x_i)\mathscr{B}$ if and only if there is a sequence s' that differs from s in at most the ith component such that s' satisfies $\mathscr{B}$. (In other words $s = (s_1, s_2, \ldots, s_i, \ldots)$ satisfies $(\exists x_i)\mathscr{B}$ if and only if there is an element c in the domain D such that the sequence $(s_1, s_2, \ldots, c, \ldots)$ satisfies $\mathscr{B}$.)

(VI) $\models_M \mathscr{B}$ if and only if $\models_M (\forall x_i)\mathscr{B}$. We can extend this result in the following way. By the *closure*‡ of $\mathscr{B}$ we mean the closed wf obtained from $\mathscr{B}$ by prefixing in universal quantifiers those variables, in order of descending subscripts, that are free in $\mathscr{B}$. If $\mathscr{B}$ has no free variables, the closure of $\mathscr{B}$ is defined to be $\mathscr{B}$ itself. For example, if $\mathscr{B}$ is $A_1^2(x_2, x_5) \Rightarrow \neg(\exists x_2)A_1^3(x_1, x_2, x_3)$, its closure is $(\forall x_5)(\forall x_3)(\forall x_2)(\forall x_1)\mathscr{B}$. It follows from (VI) that a wf $\mathscr{B}$ is true if and only if its closure is true.

(VII) Every instance of a tautology is true for any interpretation. (An *instance* of a statement form is a wf obtained from the statement form by substituting wfs for all statement letters, with all occurrences of the same statement letter being replaced by the same wf. Thus, an instance of $A_1 \Rightarrow \neg A_2 \vee A_1$ is $A_1^1(x_2) \Rightarrow (\neg(\forall x_1)A_1^1(x_1)) \vee A_1^1(x_2)$.) To prove (VII), show that all instances of the axioms of the system L are true and then use (III) and Proposition 1.14.

(VIII) If the free variables (if any) of a wf $\mathscr{B}$ occur in the list $x_{i_1}, \ldots, x_{i_k}$ and if the sequences s and s' have the same components in the i_1th, …, i_kth places, then s satisfies $\mathscr{B}$ if and only if s' satisfies $\mathscr{B}$ [*Hint*: Use induction on the number of connectives and quantifiers in $\mathscr{B}$. First prove this lemma: If the variables in a term t occur in the list $x_{i_1}, \ldots, x_{i_k}$, and if s and s' have the same components in the

†Remember that $\mathscr{B} \wedge \mathscr{C}, \mathscr{B} \vee \mathscr{C}, \mathscr{B} \Leftrightarrow \mathscr{C}$ and $(\exists x_i)\mathscr{B}$ are abbreviations for $\neg(\mathscr{B} \Rightarrow \neg\mathscr{C}), \neg\mathscr{B} \Rightarrow \mathscr{C}, (\mathscr{B} \Rightarrow \mathscr{C}) \wedge (\mathscr{C} \Rightarrow \mathscr{B})$ and $\neg(\forall x_i)\neg\mathscr{B}$, respectively.

‡A better term for *closure* would be *universal closure*.

i_1th, ..., i_kth places, then $s^*(t) = (s')^*(t)$. In particular, if t contains no variables at all, $s^*(t) = (s')^*(t)$ for any sequences s and s'.]

Although, by (VIII), a particular wf $\mathscr{B}$ with k free variables is essentially satisfied or not only by k-tuples, rather than by denumerable sequences, it is more convenient for a general treatment of satisfaction to deal with infinite rather than finite sequences. If we were to define satisfaction using finite sequences, conditions 3 and 4 of the definition of satisfaction would become much more complicated.

Let $x_{i_1}, \ldots, x_{i_k}$ be k distinct variables in order of increasing subscripts. Let $\mathscr{B}(x_{i_1}, \ldots, x_{i_k})$ be a wf that has $x_{i_1}, \ldots, x_{i_k}$ as its only free variables. The set of k-tuples $\langle b_1, \cdots, b_k \rangle$ of elements of the domain D such that any sequence with $b_1, \cdots, b_k$ in its i_1th, ..., i_kth places, respectively, satisfies $\mathscr{B}(x_{i_1}, \ldots, x_{i_k})$ is called the *relation* (or *property*[†]) *of the interpretation defined by* $\mathscr{B}$. Extending our terminology, we shall say that every k-tuple $\langle b_1, \ldots, b_k \rangle$ in this relation *satisfies* $\mathscr{B}(x_{i_1}, \ldots, x_{i_k})$ in the interpretation M; this will be written $\models_M \mathscr{B}[b_1, \ldots, b_k]$. This extended notion of satisfaction corresponds to the original intuitive notion.

Examples

1. If the domain D of M is the set of human beings, $A_1^2(x, y)$ is interpreted as x *is a brother of* y, and $A_2^2(x, y)$ is interpreted as x *is a parent of* y, then the binary relation on D corresponding to the wf $\mathscr{B}(x_1, x_2)$: $(\exists x_3)(A_1^2(x_1, x_3) \wedge A_2^2(x_3, x_2))$ is the relation of unclehood. $\models_M \mathscr{B}[b, c]$ when and only when b is an uncle of c.
2. If the domain is the set of positive integers, A_1^2 is interpreted as $=$, f_1^2 is interpreted as multiplication, and a_1 is interpreted as 1, then the wf $\mathscr{B}(x_1)$:

$$\neg A_1^2(x_1, a_1) \wedge (\forall x_2)((\exists x_3) A_1^2(x_1, f_1^2(x_2, x_3)) \Rightarrow A_1^2(x_2, x_1) \vee A_1^2(x_2, a_1))$$

determines the property of being a prime number. Thus $\models_M \mathscr{B}[k]$ if and only if k is a prime number.

(IX) If $\mathscr{B}$ is a closed wf of a language $\mathscr{L}$, then, for any interpretation M, either $\models_M \mathscr{B}$ or $\models_M \neg\mathscr{B}$ – that is, either $\mathscr{B}$ is true for M or $\mathscr{B}$ is false for M. [*Hint:* Use (VIII).] Of course, $\mathscr{B}$ may be true for some interpretations and false for others. (As an example, consider $A_1^1(a_1)$. If M is an interpretation whose domain is the set of positive integers, A_1^1 is interpreted as the property of being a prime, and the interpretation of a_1 is 2, then $A_1^1(a_1)$ is true. If we change the interpretation by interpreting a_1 as 4, then $A_1^1(a_1)$ becomes false.)

If $\mathscr{B}$ is not closed – that is, if $\mathscr{B}$ contains free variables – $\mathscr{B}$ may be neither true nor false for some interpretation. For example, if $\mathscr{B}$ is $A_1^2(x_1, x_2)$ and we consider an interpretation in which the domain is the set of integers and

[†]A property is defined when $k = 1$.

$A_1^2(y, z)$ is interpreted as $y < z$, then $\mathscr{B}$ is satisfied by only those sequences $s = (s_1, s_2, \ldots)$ of integers in which $s_1 < s_2$. Hence, $\mathscr{B}$ is neither true nor false for this interpretation. On the other hand, there are wfs that are not closed but that nevertheless are true or false for every interpretation. A simple example is the wf $A_1^1(x_1) \vee \neg A_1^1(x_1)$, which is true for every interpretation.

(X) Assume t is free for x_i in $\mathscr{B}(x_i)$. Then $(\forall x_i)\mathscr{B}(x_i) \Rightarrow \mathscr{B}(t)$ is true for all interpretations.

The proof of (X) is based upon the following lemmas.

LEMMA 1

If t and u are terms, s is a sequence in Σ, t' results from t by replacing all occurrences of x_i by u, and s' results from s by replacing the ith component of s by $s^*(u)$, then $s^*(t') = (s')^*(t)$. [*Hint:* Use induction on the length of t.†]

LEMMA 2

Let t be free for x_i in $\mathscr{B}(x_i)$. Then:

(a) A sequence $s = (s_1, s_2, \ldots)$ satisfies $\mathscr{B}(t)$ if and only if the sequence s', obtained from s by substituting $s^*(t)$ for s_i in the ith place, satisfies $\mathscr{B}(x_i)$. [*Hint:* Use induction on the number of occurrences of connectives and quantifiers in $\mathscr{B}(x_i)$, applying Lemma 1.]

(b) If $(\forall x_i)\mathscr{B}(x_i)$ is satisfied by the sequence s, then $\mathscr{B}(t)$ also is satisfied by s.

(XI) If $\mathscr{B}$ does not contain x_i free, then $(\forall x_i)(\mathscr{B} \Rightarrow \mathscr{C}) \Rightarrow (\mathscr{B} \Rightarrow (\forall x_i)\mathscr{C})$ is true for all interpretations.

Proof

Assume (XI) is not correct. Then $(\forall x_i)(\mathscr{B} \Rightarrow \mathscr{C}) \Rightarrow (\mathscr{B} \Rightarrow (\forall x_i)\mathscr{C})$ is not true for some interpretation. By condition 3 of the definition of satisfication, there is a sequence s such that s satisfies $(\forall x_i)(\mathscr{B} \Rightarrow \mathscr{C})$ and s does not satisfy $\mathscr{B} \Rightarrow (\forall x_i)\mathscr{C}$. From the latter and condition 3, s satisfies $\mathscr{B}$ and s does not satisfy $(\forall x_i)\mathscr{C}$. Hence, by condition 4, there is a sequence s', differing from s in at most the ith place, such that s' does not satisfy $\mathscr{C}$. Since x_i is free in neither $(\forall x_i)(\mathscr{B} \Rightarrow \mathscr{C})$ nor $\mathscr{B}$, and since s satisfies both of these wfs, it follows by (VIII) that s' also satisfies both $(\forall x_i)(\mathscr{B} \Rightarrow \mathscr{C})$ and $\mathscr{B}$. Since s' satisfies

†The *length* of an expression is the number of occurrences of symbols in the expression.

$(\forall x_i)(\mathscr{B} \Rightarrow \mathscr{C})$, it follows by condition 4 that s' satisfies $\mathscr{B} \Rightarrow \mathscr{C}$. Since s' satisfies $\mathscr{B} \Rightarrow \mathscr{C}$ and $\mathscr{B}$, condition 3 implies that s' satisfies $\mathscr{C}$, which contradicts the fact that s' does not satisfy $\mathscr{C}$. Hence, (XI) is established.

Exercises

2.12 Verify (I) -(X).

2.13 Prove that a closed wf $\mathscr{B}$ is true for M if and only if $\mathscr{B}$ is satisfied by *some* sequence s in Σ. (Remember that Σ is the set of denumerable sequences of elements in the domain of M.)

2.14 Find the properties or relations determined by the following wfs and interpretations.

(a) $[(\exists u)A_1^2(f_1^2(x,u),y)] \wedge [(\exists v)A_1^2(f_1^2(x,v),z)]$, where the domain D is the set of integers, A_1^2 is $=$, and f_1^2 is multiplication.

(b) Here, D is the set of non-negative integers, A_1^2 is $=$, a_1 denotes 0, f_1^2 is addition, and f_2^2 is multiplication.
   (i) $[(\exists z)(\neg A_1^2(z,a_1) \wedge A_1^2(f_1^2(x,z),y))]$
   (ii) $(\exists y)A_1^2(x,f_2^2(y,y))$

(c) $(\exists x_3)A_1^2(f_1^2(x_1,x_3),x_2)$, where D is the set of positive integers, A_1^2 is $=$, and f_1^2 is multiplication,

(d) $A_1^1(x_1) \wedge (\forall x_2)\neg A_1^2(x_1,x_2)$, where D is the set of all living people, $A_1^1(x)$ means x *is a man* and $A_1^2(x,y)$ means x *is married to* y.

(e) (i) $(\exists x_1)(\exists x_2)(A_1^2(x_1,x_3) \wedge A_1^2(x_2,x_4) \wedge A_2^2(x_1,x_2))$
   (ii) $(\exists x_3)(A_1^2(x_1,x_3) \wedge A_1^2(x_3,x_2))$
   where D is the set of all people, $A_1^2(x,y)$ means x *is a parent of* y, and $A_2^2(x,y)$ means x *and* y *are siblings.*

(f) $(\forall x_3)((\exists x_4)(A_1^2(f_1^2(x_4,x_3),x_1) \wedge (\exists x_4)(A_1^2(f_1^2(x_4,x_3),x_2)) \Rightarrow A_1^2(x_3,a_1))$, where D is the set of positive integers, A_1^2 is $=$, f_1^2 is multiplication, and a_1 denotes 1.

(g) $\neg A_1^2(x_2,x_1) \wedge (\exists y)(A_1^2(y,x_1) \wedge A_2^2(x_2,y))$,
   where D is the set of all people, $A_1^2(u,v)$ means u *is a parent of* v, and $A_2^2(u,v)$ means u *is a wife of* v.

2.15 For each of the following sentences and interpretations, write a translation into ordinary English and determine its truth or falsity.

(a) The domain D is the set of non-negative integers, A_1^2 is $=$, f_1^2 is addition, f_2^2 is multiplication, a_1 denotes 0 and a_2 denotes 1.
   (i) $(\forall x)(\exists y)(A_1^2(x,f_1^2(y,y)) \vee A_1^2(x,f_1^2(f_1^2(y,y),a_2)))$
   (ii) $(\forall x)(\forall y)(A_1^2(f_2^2(x,y),a_1) \Rightarrow A_1^2(x,a_1) \vee A_1^2(y,a_1))$
   (iii) $(\exists y)A_1^2(f_1^2(y,y),a_2)$

(b) Here, D is the set of integers, A_1^2 is $=$, and f_1^2 is addition.
   (i) $(\forall x_1)(\forall x_2)A_1^2(f_1^2(x_1,x_2),f_1^2(x_2,x_1))$
   (ii) $(\forall x_1)(\forall x_2)(\forall x_3)A_1^2(f_1^2(x_1,f_1^2(x_2,x_3)),f_1^2(f_1^2(x_1,x_2),x_3))$
   (iii) $(\forall x_1)(\forall x_2)(\exists x_3)A_1^2(f_1^2(x_1,x_3),x_2)$

(c) The wfs are the same as in part (b), but the domain is the set of positive integers, A_1^2 is $=$, and $f_1^2(x,y)$ is x^y.

(d) The domain is the set of rational numbers, A_1^2 is $=$, A_2^2 is $<$, f_1^2 is multiplication, $f_1^1(x)$ is $x+1$, and a_1 denotes 0.
(i) $(\exists x)A_1^2(f_1^2(x,x),f_1^1(f_1^1(a_1)))$
(ii) $(\forall x)(\forall y)(A_2^2(x,y) \Rightarrow (\exists z)(A_2^2(x,z) \wedge A_2^2(z,y)))$
(iii) $(\forall x)(\neg A_1^2(x,a_1) \Rightarrow (\exists y)A_1^2(f_1^2(x,y),f_1^1(a_1)))$

(e) The domain is the set of non-negative integers, $A_1^2(u,v)$ means $u \leqslant v$, and $A_1^3(u,v,w)$ means $u+v=w$.
(i) $(\forall x)(\forall y)(\forall z)(A_1^3(x,y,z) \Rightarrow A_1^3(y,x,z))$
(ii) $(\forall x)(\forall y)(A_1^3(x,x,y) \Rightarrow A_1^2(x,y))$
(iii) $(\forall x)(\forall y)(A_1^2(x,y) \Rightarrow A_1^3(x,x,y))$
(iv) $(\exists x)(\forall y)A_1^3(x,y,y)$
(v) $(\exists y)(\forall x)A_1^2(x,y)$
(vi) $(\forall x)(\forall y)(A_1^2(x,y) \Leftrightarrow (\exists z)A_1^3(x,z,y))$

(f) The domain is the set of natural numbers, $A_1^2(u,v)$ means $u=v$, $f_1^2(u,v)=u+v$, and $f_2^2(u,v)=u \cdot v$
$(\forall x)(\exists y)(\exists z)A_1^2(x,f_1^2(f_2^2(y,y),f_2^2(z,z)))$

DEFINITIONS

A wf $\mathscr{B}$ is said to be *logically valid* if and only if $\mathscr{B}$ is true for every interpretation.†

$\mathscr{B}$ is said to be *satisfiable* if and only if there is an interpretation for which $\mathscr{B}$ is satisfied by at least one sequence.

It is obvious that $\mathscr{B}$ is logically valid if and only if $\neg\mathscr{B}$ is not satisfiable, and $\mathscr{B}$ is satisfiable if and only if $\neg\mathscr{B}$ is not logically valid.

If $\mathscr{B}$ is a closed wf, then we know that $\mathscr{B}$ is either true or false for any given interpretation; that is, $\mathscr{B}$ is satisfied by all sequences or by none. Therefore, if $\mathscr{B}$ is closed, then $\mathscr{B}$ is satisfiable if and only if $\mathscr{B}$ is true for some interpretation.

A set Γ of wfs is said to be *satisfiable* if and only if there is an interpretation in which there is a sequence that satisfies every wf of Γ.

It is impossible for both a wf $\mathscr{B}$ and its negation $\neg\mathscr{B}$ to be logically valid. For if $\mathscr{B}$ is true for an interpretation, then $\neg\mathscr{B}$ is false for that interpretation.

We say that $\mathscr{B}$ is *contradictory* if and only if $\mathscr{B}$ is false for every interpretation, or, equivalently, if and only if $\neg\mathscr{B}$ is logically valid.

$\mathscr{B}$ is said to *logically imply* $\mathscr{C}$ if and only if, in every interpretation, every sequence that satisfies $\mathscr{B}$ also satisfies $\mathscr{C}$. More generally, $\mathscr{C}$ is said to be a

†The mathematician and philosopher G.W. Leibniz (1646–1716) gave a similar definition: $\mathscr{B}$ is logically valid if and only if $\mathscr{B}$ is true in all 'possible worlds'.

logical consequence of a set Γ of wfs if and only if, in every interpretation, every sequence that satisfies every wf in Γ also satisfies $\mathscr{C}$.

$\mathscr{B}$ and $\mathscr{C}$ are said to be *logically equivalent* if and only if they logically imply each other.

The following assertions are easy consequences of these definitions.

1. $\mathscr{B}$ logically implies $\mathscr{C}$ if and only if $\mathscr{B} \Rightarrow \mathscr{C}$ is logically valid.
2. $\mathscr{B}$ and $\mathscr{C}$ are logically equivalent if and only if $\mathscr{B} \Leftrightarrow \mathscr{C}$ is logically valid.
3. If $\mathscr{B}$ logically implies $\mathscr{C}$ and $\mathscr{B}$ is true in a given interpretation, then so is $\mathscr{C}$.
4. If $\mathscr{C}$ is a logical consequence of a set Γ of wfs and all wfs in Γ are true in a given interpretation, then so is $\mathscr{C}$.

Exercise 2.16

Prove assertions 1–4.

Examples

1. Every instance of a tautology is logically valid (VII).
2. If t is free for x in $\mathscr{B}(x)$, then $(\forall x)\mathscr{B}(x) \Rightarrow \mathscr{B}(t)$ is logically valid (X).
3. If $\mathscr{B}$ does not contain x free, then $(\forall x)(\mathscr{B} \Rightarrow \mathscr{C}) \Rightarrow (\mathscr{B} \Rightarrow (\forall x)\mathscr{C})$ is logically valid (XI).
4. $\mathscr{B}$ is logically valid if and only if $(\forall y_1) \ldots (\forall y_n)\mathscr{B}$ is logically valid (VI).
5. The wf $(\forall x_2)(\exists x_1)A_1^2(x_1, x_2) \Rightarrow (\exists x_1)(\forall x_2)A_1^2(x_1, x_2)$ is not logically valid. As a counterexample, let the domain D be the set of integers and let $A_1^2(y, z)$ mean $y < z$. Then $(\forall x_2)(\exists x_1)A_1^2(x_1, x_2)$ is true but $(\exists x_1)(\forall x_2)$ $A_1^2(x_1, x_2)$ is false.

Exercises

2.17 Show that the following wfs are not logically valid.

(a) $[(\forall x_1)A_1^1(x_1) \Rightarrow (\forall x_1)A_2^1(x_1)] \Rightarrow [(\forall x_1)(A_1^1(x_1) \Rightarrow A_2^1(x_1))]$

(b) $[(\forall x_1)(A_1^1(x_1) \vee A_2^1(x_1))] \Rightarrow [((\forall x_1)A_1^1(x_1)) \vee (\forall x_1)A_2^1(x_1)]$

2.18 Show that the following wfs are logically valid.†

(a) $\mathscr{B}(t) \Rightarrow (\exists x_i)\mathscr{B}(x_i)$ if t is free for x_i in $\mathscr{B}(x_i)$

(b) $(\forall x_i)\mathscr{B} \Rightarrow (\exists x_i)\mathscr{B}$

(c) $(\forall x_i)(\forall x_j)\mathscr{B} \Rightarrow (\forall x_j)(\forall x_i)\mathscr{B}$

(d) $(\forall x_i)\mathscr{B} \Leftrightarrow \neg(\exists x_i)\neg\mathscr{B}$

(e) $(\forall x_i)(\mathscr{B} \Rightarrow \mathscr{C}) \Rightarrow ((\forall x_i)\mathscr{B} \Rightarrow (\forall x_i)\mathscr{C})$

(f) $((\forall x_i)\mathscr{B}) \wedge (\forall x_i)\mathscr{C} \Leftrightarrow (\forall x_i)(\mathscr{B} \wedge \mathscr{C})$

(g) $((\forall x_i)\mathscr{B}) \vee (\forall x_i)\mathscr{C} \Rightarrow (\forall x_i)(\mathscr{B} \vee \mathscr{C})$

(h) $(\exists x_i)(\exists x_j)\mathscr{B} \Leftrightarrow (\exists x_j)(\exists x_i)\mathscr{B}$

†At this point, one can use intuitive arguments or one can use the rigorous definitions of satisfaction and truth, as in the argument above for (XI). Later on, we shall discover another method for showing logical validity.

(i) $(\exists x_i)(\forall x_j)\mathscr{B} \Rightarrow (\forall x_j)(\exists x_i)\mathscr{B}$

2.19 (a) If $\mathscr{B}$ is a closed wf, show that $\mathscr{B}$ logically implies $\mathscr{C}$ if and only if $\mathscr{C}$ is true for every interpretation for which $\mathscr{B}$ is true.

(b) Although, by (VI), $(\forall x_1)A_1^1(x_1)$ is true whenever $A_1^1(x_1)$ is true, find an interpretation for which $A_1^1(x_1) \Rightarrow (\forall x_1)A_1^1(x_1)$ is not true. (Hence, the hypothesis that $\mathscr{B}$ is a closed wf is essential in (a).)

2.20 Prove that, if the free variables of $\mathscr{B}$ are $y_1, \ldots, y_n$, then $\mathscr{B}$ is satisfiable if and only if $(\exists y_1)\ldots(\exists y_n)\mathscr{B}$ is satisfiable.

2.21 Produce counterexamples to show that the following wfs are not logically valid (that is, in each case, find an interpretation for which the wf is not true).

(a) $[(\forall x)(\forall y)(\forall z)(A_1^2(x,y) \wedge A_1^2(y,z) \Rightarrow A_1^2(x,z)) \wedge (\forall x)\neg A_1^2(x,x)] \Rightarrow (\exists x)(\forall y)\neg A_1^2(x,y)$

(b) $(\forall x)(\exists y)A_1^2(x,y) \Rightarrow (\exists y)A_1^2(y,y)$

(c) $(\exists x)(\exists y)A_1^2(x,y) \Rightarrow (\exists y)A_1^2(y,y)$

(d) $[(\exists x)A_1^1(x) \Leftrightarrow (\exists x)A_2^1(x)] \Rightarrow (\forall x)(A_1^1(x) \Leftrightarrow A_2^1(x))$

(e) $(\exists x)(A_1^1(x) \Rightarrow A_2^1(x)) \Rightarrow ((\exists x)A_1^1(x) \Rightarrow (\exists x)A_2^1(x))$

(f) $[(\forall x)(\forall y)(A_1^2(x,y) \Rightarrow A_1^2(y,x)) \wedge (\forall x)(\forall y)(\forall z)(A_1^2(x,y) \wedge A_1^2(y,z) \Rightarrow A_1^2(x,z))] \Rightarrow (\forall x)A_1^2(x,x)$

(g)$^{\mathrm{D}}$ $(\exists x)(\forall y)(A_1^2(x,y) \wedge \neg A_1^2(y,x) \Rightarrow [A_1^2(x,x) \Leftrightarrow A_1^2(y,y)])$

(h) $(\forall x)(\forall y)(\forall z)(A_1^2(x,x) \wedge (A_1^2(x,z) \Rightarrow A_1^2(x,y) \vee A_1^2(y,z))) \Rightarrow (\exists y)(\forall z)A_1^2(y,z)$

(i) $(\exists x)(\forall y)(\exists z)((A_1^2(y,z) \Rightarrow A_1^2(x,z)) \Rightarrow (A_1^2(x,x) \Rightarrow A_1^2(y,x)))$

2.22 By introducing appropriate notation, write the sentences of each of the following arguments as wfs and determine whether the argument is correct, that is, determine whether the conclusion is logically implied by the conjunction of the premisses

(a) All scientists are neurotic. No vegetarians are neurotic. Therefore, no vegetarians are scientists.

(b) All men are animals. Some animals are carnivorous. Therefore, some men are carnivorous.

(c) Some geniuses are celibate. Some students are not celibate. Therefore, some students are not geniuses.

(d) Any barber in Jonesville shaves exactly those men in Jonesville who do not shave themselves. Hence, there is no barber in Jonesville.

(e) For any numbers x, y, z, if $x > y$ and $y > z$, then $x > z$. $x > x$ is false for all numbers x. Therefore, for any numbers x and y, if $x > y$, then it is not the case that $y > x$.

(f) No student in the statistics class is smarter than every student in the logic class. Hence, some student in the logic class is smarter than every student in the statistics class.

(g) Everyone who is sane can understand mathematics. None of Hegel's sons can understand mathematics. No madmen are fit to vote. Hence, none of Hegel's sons is fit to vote.

(h) For every set x, there is a set y such that the cardinality of y is greater than the cardinality of x. If x is included in y, the cardinality of x is not greater than the cardinality of y. Every set is included in V. Hence, V is not a set.

(i) For all positive integers $x, x \leqslant x$. For all positive integers x, y, z, if $x \leqslant y$ and $y \leqslant z$, then $x \leqslant z$. For all positive integers x and y, $x \leqslant y$ or $y \leqslant x$. Therefore, there is a positive integer y such that, for all positive integers $x, y \leqslant x$.

(j) For any integers x, y, z, if $x > y$ and $y > z$, then $x > z$. $x > x$ is false for all integers x. Therefore, for any integers x and y, if $x > y$, then it is not the case that $y > x$.

2.23 Determine whether the following sets of wfs are compatible – that is, whether their conjunction is satisfiable.

(a) $(\exists x)(\exists y)A_1^2(x,y)$
$(\forall x)(\forall y)(\exists z)(A_1^2(x,z) \wedge A_1^2(z,y))$

(b) $(\forall x)(\exists y)A_1^2(y,x)$
$(\forall x)(\forall y)(A_1^2(x,y) \Rightarrow \neg A_1^2(y,x))$
$(\forall x)(\forall y)(\forall z)(A_1^2(x,y) \wedge A_1^2(y,z) \Rightarrow A_1^2(x,z))$

(c) All unicorns are animals.
No unicorns are animals.

2.24 Determine whether the following wfs are logically valid.

(a) $\neg(\exists y)(\forall x)(A_1^2(x,y) \Leftrightarrow \neg A_1^2(x,x))$

(b) $[(\exists x)A_1^1(x) \Rightarrow (\exists x)A_2^1(x)] \Rightarrow (\exists x)(A_1^1(x) \Rightarrow A_2^1(x))$

(c) $(\exists x)(A_1^1(x) \Rightarrow (\forall y)A_1^1(y))$

(d) $(\forall x)(A_1^1(x) \vee A_2^1(x)) \Rightarrow (((\forall x)A_1^1(x)) \vee (\exists x)A_2^1(x))$

(e) $(\exists x)(\exists y)(A_1^2(x,y) \Rightarrow (\forall z)A_1^2(z,y))$

(f) $(\exists x)(\exists y)(A_1^1(x) \Rightarrow A_2^1(y)) \Rightarrow (\exists x)(A_1^1(x) \Rightarrow A_2^1(x))$

(g) $(\forall x)(A_1^1(x) \Rightarrow A_2^1(x)) \Rightarrow \neg(\forall x)(A_1^1(x) \Rightarrow \neg A_2^1(x))$

(h) $(\exists x)A_1^2(x,x) \Rightarrow (\exists x)(\exists y)A_1^2(x,y)$

2.25 Exhibit a logically valid wf that is not an instance of a tautology. However, show that any logically valid *open* wf (that is, a wf without quantifiers) must be an instance of a tautology.

2.26 (a) Find a satisfiable closed wf that is not true in any interpretation whose domain has only one member.

(b) Find a satisfiable closed wf that is not true in any interpretation whose domain has fewer than three members.

2.3 FIRST-ORDER THEORIES

In the case of the propositional calculus, the method of truth tables provides an effective test as to whether any given statement form is a tautology. However, there does not seem to be any effective process for determining whether a given wf is logically valid, since, in general, one has to check the truth of a wf for interpretations with arbitrarily large finite or infinite domains. In fact, we shall see later that, according to a plausible definition of 'effective', it may actually be proved that there is no effective way to test for logical validity. The axiomatic method, which was a luxury in the study of the propositional calculus, thus appears to be a necessity in the study of wfs involving quantifiers,[†] and we therefore turn now to the consideration of first-order theories.

Let $\mathscr{L}$ be a first-order language. A *first-order theory* in the language $\mathscr{L}$ will be a formal theory K whose symbols and wfs are the symbols and wfs of $\mathscr{L}$ and whose axioms and rules of inference are specified in the following way.[‡]

The axioms of K are divided into two classes: the logical axioms and the proper (or non-logical) axioms.

LOGICAL AXIOMS

If $\mathscr{B}, \mathscr{C}$ and $\mathscr{D}$ are wfs of $\mathscr{L}$, then the following are logical axioms of K:

(A1) $\mathscr{B} \Rightarrow (\mathscr{C} \Rightarrow \mathscr{B})$

(A2) $(\mathscr{B} \Rightarrow (\mathscr{C} \Rightarrow \mathscr{D})) \Rightarrow ((\mathscr{B} \Rightarrow \mathscr{C}) \Rightarrow (\mathscr{B} \Rightarrow \mathscr{D}))$

(A3) $(\neg\mathscr{C} \Rightarrow \neg\mathscr{B}) \Rightarrow ((\neg\mathscr{C} \Rightarrow \mathscr{B}) \Rightarrow \mathscr{C})$

(A4) $(\forall x_i)\mathscr{B}(x_i) \Rightarrow \mathscr{B}(t)$ if $\mathscr{B}(x_i)$ is a wf of $\mathscr{L}$ and t is a term of $\mathscr{L}$ that is free for x_i in $\mathscr{B}(x_i)$. Note here that t may be identical with x_i so that all wfs $(\forall x_i)\mathscr{B} \Rightarrow \mathscr{B}$ are axioms by virtue of axiom (A4).

(A5) $(\forall x_i)(\mathscr{B} \Rightarrow \mathscr{C}) \Rightarrow (\mathscr{B} \Rightarrow (\forall x_i)\mathscr{C})$ if $\mathscr{B}$ contains no free occurrences of x_i.

[†]There is still another reason for a formal axiomatic approach. Concepts and propositions that involve the notion of interpretation and related ideas such as truth and model are often called *semantical* to distinguish them from *syntactical* concepts, which refer to simple relations among symbols and expressions of precise formal languages. Since semantical notions are set-theoretic in character, and since set theory, because of the paradoxes, is considered a rather shaky foundation for the study of mathematical logic, many logicians consider a syntactical approach, consisting of a study of formal axiomatic theories using only rather weak number-theoretic methods, to be much safer. For further discussions, see the pioneering study on semantics by Tarski (1936), as well as Kleene (1952), Church (1956) and Hilbert and Bernays (1934).

[‡]The reader might wish to review the definition of *formal theory* in Section 1.4. We shall use the terminology (proof, theorem, consequence, axiomatic, $\vdash \mathscr{B}$ etc.) and notation ($\Gamma \vdash \mathscr{B}, \vdash \mathscr{B}$) introduced there.

PROPER AXIOMS

These cannot be specified, since they vary from theory to theory. A first-order theory in which there are no proper axioms is called a first-order *predicate calculus.*

RULES OF INFERENCE

The rules of inference of any first-order theory are:

1. Modus ponens: $\mathscr{C}$ follows from $\mathscr{B}$ and $\mathscr{B} \Rightarrow \mathscr{C}$.
2. Generalization: $(\forall x_i)\mathscr{B}$ follows from $\mathscr{B}$.

We shall use the abbreviations MP and Gen, respectively, to indicate applications of these rules.

DEFINITION

Let K be a first-order theory in the language $\mathscr{L}$. By a *model* of K we mean an interpretation of $\mathscr{L}$ for which all the axioms of K are true.

By (III) and (VI) on page 61, if the rules of modus ponens and generalization are applied to wfs that are true for a given interpretation, then the results of these applications are also true. Hence *every theorem of* K *is true in every model of* K.

As we shall see, the logical axioms are so designed that the logical consequences (in the sense defined on pages 65–6) of the closures of the axioms of K are precisely the theorems of K. In particular, if K is a first-order predicate calculus, it turns out that the theorems of K are just those wfs of K that are logically valid.

Some explanation is needed for the restrictions in axiom schemas (A4) and (A5). In the case of (A4), if t were not free for x_i in $\mathscr{B}(x_i)$, the following unpleasant result would arise: let $\mathscr{B}(x_1)$ be $\neg(\forall x_2)A_1^2(x_1, x_2)$ and let t be x_2. Notice that t is not free for x_1 in $\mathscr{B}(x_1)$. Consider the following pseudo-instance of axiom (A4):

$$(\nabla) \quad (\forall x_1)(\neg(\forall x_2)A_1^2(x_1, x_2)) \Rightarrow \neg(\forall x_2)A_1^2(x_2, x_2)$$

Now take as interpretation any domain with at least two members and let A_1^2 stand for the identity relation. Then the antecedent of (∇) is true and the consequent false. Thus, (∇) is false for this interpretation.

In the case of axiom (A5), relaxation of the restriction that x_i not be free in $\mathscr{B}$ would lead to the following disaster. Let $\mathscr{B}$ and $\mathscr{C}$ both be $A_1^1(x_1)$. Thus, x_1 is free in $\mathscr{B}$. Consider the following pseudo-instance of axiom (A5):

$$(\nabla\nabla) \quad (\forall x_1)(A_1^1(x_1) \Rightarrow A_1^1(x_1)) \Rightarrow (A_1^1(x_1) \Rightarrow (\forall x_1)A_1^1(x_1))$$

The antecedent of $(\nabla\nabla)$ is logically valid. Now take as domain the set of integers and let $A_1^1(x)$ mean that x *is even*. Then $(\forall x_1)A_1^1(x_1)$ is false. So, any sequence $s = (s_1, s_2, \ldots)$ for which s_1 is even does not satisfy the consequent of $(\nabla\nabla)$.† Hence, $(\nabla\nabla)$ is not true for this interpretation.

Examples of first-order theories

1. *Partial order.* Let the language $\mathscr{L}$ have a single predicate letter A_2^2 and no function letters and individual constants. We shall write $x_i < x_j$ instead of $A_2^2(x_i, x_j)$. The theory K has two proper axioms.
   (a) $(\forall x_1)(\neg\, x_1 < x_1)$ (irreflexivity)
   (b) $(\forall x_1)(\forall x_2)(\forall x_3)(x_1 < x_2 \wedge x_2 < x_3 \Rightarrow x_1 < x_3)$ (transitivity)

   A model of the theory is called a *partially ordered structure.*

2. *Group theory.* Let the language $\mathscr{L}$ have one predicate letter A_1^2, one function letter f_1^2, and one individual constant a_1. To conform with ordinary notation, we shall write $t = s$ instead of $A_1^2(t, s)$, $t + s$ instead of $f_1^2(t, s)$, and 0 instead of a_1. The proper axioms of K are:

   (a) $(\forall x_1)(\forall x_2)(\forall x_3)(x_1 + (x_2 + x_3) = (x_1 + x_2) + x_3)$ (associativity)
   (b) $(\forall x_1)(0 + x_1 = x_1)$ (identity)
   (c) $(\forall x_1)(\exists x_2)(x_2 + x_1 = 0)$ (inverse)
   (d) $(\forall x_1)(x_1 = x_1)$ (reflexivity of $=$)
   (e) $(\forall x_1)(\forall x_2)(x_1 = x_2 \Rightarrow x_2 = x_1)$ (symmetry of $=$)
   (f) $(\forall x_1)(\forall x_2)(\forall x_3)(x_1 = x_2 \wedge x_2 = x_3 \Rightarrow x_1 = x_3)$ (transitivity of $=$)
   (g) $(\forall x_1)(\forall x_2)(\forall x_3)(x_2 = x_3 \Rightarrow$
   $x_1 + x_2 = x_1 + x_3 \wedge x_2 + x_1 = x_3 + x_1)$ (substitutivity of $=$)

A model for this theory, in which the interpretation of $=$ is the identity relation, is called a *group*. A group is said to be *abelian* if, in addition, the wf $(\forall x_1)(\forall x_2)(x_1 + x_2 = x_2 + x_1)$ is true.

The theories of partial order and of groups are both axiomatic. In general, any theory with a finite number of proper axioms is axiomatic, since it is obvious that one can effectively decide whether any given wf is a logical axiom.

2.4 PROPERTIES OF FIRST-ORDER THEORIES

All the results in this section refer to an arbitrary first-order theory K. Instead of writing $\vdash_K \mathscr{B}$, we shall sometimes simply write $\vdash \mathscr{B}$. Moreover, we shall refer to first-order theories simply as *theories*, unless something is said to the contrary.

†Such a sequence would satisfy $A_1^1(x_1)$, since s_1 is even, but would not satisfy $(\forall x_1)A_1^1(x_1)$, since no sequence satisfies $(\forall x_1)A_1^1(x_1)$.

PROPOSITION 2.1

Every wf $\mathscr{B}$ of K that is an instance of a tautology is a theorem of K, and it may be proved using only axioms (A1)–(A3) and MP.

Proof

$\mathscr{B}$ arises from a tautology $\mathscr{T}$ by substitution. By Proposition 1.14, there is a proof of $\mathscr{T}$ in L. In such a proof, make the same substitution of wfs of K for statement letters as were used in obtaining $\mathscr{B}$ from $\mathscr{T}$, and, for all statement letters in the proof that do not occur in $\mathscr{T}$, substitute an arbitrary wf of K. Then the resulting sequence of wfs is a proof of $\mathscr{B}$, and this proof uses only axiom schemes (A1)–(A3) and MP.

The application of Proposition 2.1 in a proof will be indicated by writing 'Tautology'.

PROPOSITION 2.2

Every theorem of a first-order predicate calculus is logically valid.

Proof

Axioms (A1)–(A3) are logically valid by property (VII) of the notion of truth (see page 61), and axioms (A4) and (A5) are logically valid by properties (X) and (XI). By properties (III) and (VI), the rules of inference MP and Gen preserve logical validity. Hence, every theorem of a predicate calculus is logically valid.

Example
The wf $(\forall x_2)(\exists x_1)A_1^2(x_1, x_2) \Rightarrow (\exists x_1)(\forall x_2)A_1^2(x_1, x_2)$ is not a theorem of any first-order predicate calculus, since it is not logically valid (by Example 5, p. 66).

DEFINITION

A theory K is *consistent* if no wf $\mathscr{B}$ and its negation $\neg\mathscr{B}$ are both provable in K. A theory is *inconsistent* if it is not consistent.

COROLLARY 2.3

Any first-order predicate calculus is consistent.

Proof

If a wf $\mathscr{B}$ and its negation $\neg\mathscr{B}$ were both theorems of a first-order predicate calculus, then, by Proposition 2.2, both $\mathscr{B}$ and $\neg\mathscr{B}$ would be logically valid, which is impossible.

Notice that, in an inconsistent theory K, every wf $\mathscr{C}$ of K is provable in K. In fact, assume that $\mathscr{B}$ and $\neg\mathscr{B}$ are both provable in K. Since the wf $\mathscr{B} \Rightarrow (\neg\mathscr{B} \Rightarrow \mathscr{C})$ is an instance of a tautology, that wf is, by Proposition 2.1, provable in K. Then two applications of MP would yield $\vdash \mathscr{C}$.

It follows from this remark that, if some wf of a theory K is not a theorem of K, then K is consistent.

The deduction theorem (Proposition 1.9) for the propositional calculus cannot be carried over without modification to first-order theories. For example, for any wf $\mathscr{B}$, $\mathscr{B} \vdash_K (\forall x_i)\mathscr{B}$, but it is not always the case that $\vdash_K \mathscr{B} \Rightarrow (\forall x_i)\mathscr{B}$. Consider a domain containing at least two elements c and d. Let K be a predicate calculus and let $\mathscr{B}$ be $A_1^1(x_1)$. Interpret A_1^1 as a property that holds only for c. Then $A_1^1(x_1)$ is satisfied by any sequence $s = (s_1, s_2, \ldots)$ in which $s_1 = c$, but $(\forall x_1)A_1^1(x_1)$ is satisfied by no sequence at all. Hence, $A_1^1(x_1) \Rightarrow (\forall x_1)A_1^1(x_1)$ is not true in this interpretation, and so it is not logically valid. Therefore, by Proposition 2.2, $A_1^1(x_1) \Rightarrow (\forall x_1)A_1^1(x_1)$ is not a theorem of K.

A modified, but still useful, form of the deduction theorem may be derived, however. Let $\mathscr{B}$ be a wf in a set Γ of wfs and assume that we are given a deduction $\mathscr{D}_1, \ldots, \mathscr{D}_n$ from Γ, together with justification for each step in the deduction. We shall say that $\mathscr{D}_i$ *depends upon* $\mathscr{B}$ in this proof if and only if:

(1) $\mathscr{D}_i$ is $\mathscr{B}$ and the justification for $\mathscr{D}_i$ is that it belongs to Γ, or
(2) $\mathscr{D}_i$ is justified as a direct consequence by MP or Gen of some preceding wfs of the sequence, where at least one of these preceding wfs depends upon $\mathscr{B}$.

Example

$\mathscr{B}, (\forall x_1)\mathscr{B} \Rightarrow \mathscr{C} \vdash (\forall x_1)\mathscr{C}$

$(\mathscr{D}_1)$	$\mathscr{B}$	Hyp
$(\mathscr{D}_2)$	$(\forall x_1)\mathscr{B}$	$(\mathscr{D}_1)$, Gen
$(\mathscr{D}_3)$	$(\forall x_1)\mathscr{B} \Rightarrow \mathscr{C}$	Hyp
$(\mathscr{D}_4)$	$\mathscr{C}$	$(\mathscr{D}_2), (\mathscr{D}_3)$, MP
$(\mathscr{D}_5)$	$(\forall x_1)\mathscr{C}$	$(\mathscr{D}_4)$, Gen

Here, $(\mathscr{D}_1)$ depends upon $\mathscr{B}$, $(\mathscr{D}_2)$ depends upon $\mathscr{B}$, $(\mathscr{D}_3)$ depends upon $(\forall x_1)\mathscr{B} \Rightarrow \mathscr{C}$, $(\mathscr{D}_4)$ depends upon $\mathscr{B}$ and $(\forall x_1)\mathscr{B} \Rightarrow \mathscr{C}$, and $(\mathscr{D}_5)$ depends upon $\mathscr{B}$ and $(\forall x_1)\mathscr{B} \Rightarrow \mathscr{C}$.

PROPOSITION 2.4

If $\mathscr{C}$ does not depend upon $\mathscr{B}$ in a deduction showing that $\Gamma, \mathscr{B} \vdash \mathscr{C}$, then $\Gamma \vdash \mathscr{C}$.

Proof

Let $\mathscr{D}_1 \cdots, \mathscr{D}_n$ be a deduction of $\mathscr{C}$ from Γ and $\mathscr{B}$, in which $\mathscr{C}$ does not depend upon $\mathscr{B}$. (In this deduction, $\mathscr{D}_n$ is $\mathscr{C}$.) As an inductive hypothesis, let us assume that the proposition is true for all deductions of length less than n. If $\mathscr{C}$ belongs to Γ or is an axiom, then $\Gamma \vdash \mathscr{C}$. If $\mathscr{C}$ is a direct consequence of one or two preceding wfs by Gen or MP, then, since $\mathscr{C}$ does not depend upon $\mathscr{B}$, neither do these preceding wfs. By the inductive hypothesis, these preceding wfs are deducible from Γ alone. Consequently, so is $\mathscr{C}$.

PROPOSITION 2.5 (DEDUCTION THEOREM)

Assume that, in some deduction showing that $\Gamma, \mathscr{B} \vdash \mathscr{C}$, no application of Gen to a wf that depends upon $\mathscr{B}$ has as its quantified variable a free variable of $\mathscr{B}$. The $\Gamma \vdash \mathscr{B} \Rightarrow \mathscr{C}$.

Proof

Let $\mathscr{D}_1, \ldots, \mathscr{D}_n$ be a deduction of $\mathscr{C}$ from Γ and $\mathscr{B}$, satisfying the assumption of our proposition. (In this deduction, $\mathscr{D}_n$ is $\mathscr{C}$.) Let us show by induction that $\Gamma \vdash \mathscr{B} \Rightarrow \mathscr{D}_i$ for each $i \leqslant n$. If $\mathscr{D}_i$ is an axiom or belongs to Γ, then $\Gamma \vdash \mathscr{B} \Rightarrow \mathscr{D}_i$, since $\mathscr{D}_i \Rightarrow (\mathscr{B} \Rightarrow \mathscr{D}_i)$ is an axiom. If $\mathscr{D}_i$ is $\mathscr{B}$, then $\Gamma \vdash \mathscr{B} \Rightarrow \mathscr{D}_i$, since, by Proposition 2.1, $\vdash \mathscr{B} \Rightarrow \mathscr{B}$. If there exist j and k less than i such that $\mathscr{D}_k$ is $\mathscr{D}_j \Rightarrow \mathscr{D}_i$, then, by inductive hypothesis, $\Gamma \vdash \mathscr{B} \Rightarrow \mathscr{D}_j$ and $\Gamma \vdash \mathscr{B} \Rightarrow (\mathscr{D}_j \Rightarrow \mathscr{D}_i)$. Now, by axiom (A2), $\vdash (\mathscr{B} \Rightarrow (\mathscr{D}_j \Rightarrow \mathscr{D}_i)) \Rightarrow ((\mathscr{B} \Rightarrow \mathscr{D}_j) \Rightarrow (\mathscr{B} \Rightarrow \mathscr{D}_i))$. Hence, by MP twice, $\Gamma \vdash \mathscr{B} \Rightarrow \mathscr{D}_i$. Finally, suppose that there is some $j < i$ such that $\mathscr{D}_i$ is $(\forall x_k)\mathscr{D}_j$. By the inductive hypothesis, $\Gamma \vdash \mathscr{B} \Rightarrow \mathscr{D}_j$, and, by the hypothesis of the theorem, either $\mathscr{D}_j$ does not depend upon $\mathscr{B}$ or x_k is not a free variable of $\mathscr{B}$. If $\mathscr{D}_j$ does not depend upon $\mathscr{B}$, then, by Proposition 2.4, $\Gamma \vdash \mathscr{D}_j$ and, consequently, by Gen, $\Gamma \vdash (\forall x_k)\mathscr{D}_j$. Thus, $\Gamma \vdash \mathscr{D}_i$. Now, by axiom (A1), $\vdash \mathscr{D}_i \Rightarrow (\mathscr{B} \Rightarrow \mathscr{D}_i)$. So, $\Gamma \vdash \mathscr{B} \Rightarrow \mathscr{D}_i$ by MP. If, on the other hand, x_k is not a free variable of $\mathscr{B}$, then, by axiom (A5), $\vdash (\forall x_k)(\mathscr{B} \Rightarrow \mathscr{D}_j) \Rightarrow (\mathscr{B} \Rightarrow (\forall x_k)\mathscr{D}_j)$. Since $\Gamma \vdash \mathscr{B} \Rightarrow \mathscr{D}_j$, we have, by Gen, $\Gamma \vdash (\forall x_k)(\mathscr{B} \Rightarrow \mathscr{D}_j)$, and so, by MP, $\Gamma \vdash \mathscr{B} \Rightarrow (\forall x_k)\mathscr{D}_j$; that is, $\Gamma \vdash \mathscr{B} \Rightarrow \mathscr{D}_i$. This completes the induction, and our proposition is just the special case $i = n$.

The hypothesis of Proposition 2.5 is rather cumbersome; the following weaker corollaries often prove to be more useful.

COROLLARY 2.6

If a deduction showing that $\Gamma, \mathscr{B} \vdash \mathscr{C}$ involves no application of Gen of which the quantified variables is free in $\mathscr{B}$, then $\Gamma \vdash \mathscr{B} \Rightarrow \mathscr{C}$.

COROLLARY 2.7

If $\mathscr{B}$ is a closed wf and $\Gamma, \mathscr{B} \vdash \mathscr{C}$, then $\Gamma \vdash \mathscr{B} \Rightarrow \mathscr{C}$.

EXTENSION OF PROPOSITIONS 2.4–2.7

In Propositions 2.4–2.7, the following additional conclusion can be drawn from the proofs. The new proof of $\Gamma \vdash \mathscr{B} \Rightarrow \mathscr{C}$ (in Proposition 2.4, of $\Gamma \vdash \mathscr{C}$) involves an application of Gen to a wf depending upon a wf $\mathscr{E}$ of Γ only if there is an application of Gen in the given proof of $\Gamma, \mathscr{B} \vdash \mathscr{C}$ that involves the same quantified variable and is applied to a wf that depends upon $\mathscr{E}$. (In the proof of Proposition 2.5, one should observe that $\mathscr{D}_j$ depends upon a premiss $\mathscr{E}$ of Γ in the original proof if and only if $\mathscr{B} \Rightarrow \mathscr{D}_j$ depends upon $\mathscr{E}$ in the new proof.)

This supplementary conclusion will be useful when we wish to apply the deduction theorem several times in a row to a given deduction – for example, to obtain $\Gamma \vdash \mathscr{D} \Rightarrow (\mathscr{B} \Rightarrow \mathscr{C})$ from $\Gamma, \mathscr{D}, \mathscr{B} \vdash \mathscr{C}$; from now on, it is to be considered an integral part of the statements of Propositions 2.4–2.7.

Example
$\vdash (\forall x_1)(\forall x_2)\mathscr{B} \Rightarrow (\forall x_2)(\forall x_1)\mathscr{B}$

Proof

1. $(\forall x_1)(\forall x_2)\mathscr{B}$	Hyp
2. $(\forall x_1)(\forall x_2)\mathscr{B} \Rightarrow (\forall x_2)\mathscr{B}$	(A4)
3. $(\forall x_2)\mathscr{B}$	1, 2, MP
4. $(\forall x_2)\mathscr{B} \Rightarrow \mathscr{B}$	(A4)
5. $\mathscr{B}$	3, 4, MP
6. $(\forall x_1)\mathscr{B}$	5, Gen
7. $(\forall x_2)(\forall x_1)\mathscr{B}$	6, Gen

Thus, by 1–7, we have $(\forall x_1)(\forall x_2)\mathscr{B} \vdash (\forall x_2)(\forall x_1)\mathscr{B}$, where, in the deduction, no application of Gen has as a quantified variable a free variable of $(\forall x_1)(\forall x_2)\mathscr{B}$. Hence, by Corollary 2.6, $\vdash (\forall x_1)(\forall x_2)\mathscr{B} \Rightarrow (\forall x_2)(\forall x_1)\mathscr{B}$.

Exercises

2.27 Derive the following theorems.
(a) $\vdash (\forall x)(\mathscr{B} \Rightarrow \mathscr{C}) \Rightarrow ((\forall x)\mathscr{B} \Rightarrow (\forall x)\mathscr{C})$

(b) $\vdash (\forall x)(\mathscr{B} \Rightarrow \mathscr{C}) \Rightarrow ((\exists x)\mathscr{B} \Rightarrow (\exists x)\mathscr{C})$
(c) $\vdash (\forall x)(\mathscr{B} \wedge \mathscr{C}) \Leftrightarrow ((\forall x)\mathscr{B}) \wedge (\forall x)\mathscr{C}$
(d) $\vdash (\forall y_1)\ldots(\forall y_n)\mathscr{B} \Rightarrow \mathscr{B}$
(e) $\vdash \neg(\forall x)\mathscr{B} \Rightarrow (\exists x)\neg\mathscr{B}$

2.28[D] Let K be a first-order theory and let $K^{\#}$ be an axiomatic theory having the following axioms:

(a) $(\forall y_1)\ldots(\forall y_n)\mathscr{B}$, where $\mathscr{B}$ is any axiom of K and $y_1, \ldots, y_n (n \geqslant 0)$ are any variables (none at all when $n = 0$);
(b) $(\forall y_1)\ldots(\forall y_n)(\mathscr{B} \Rightarrow \mathscr{C}) \Rightarrow [(\forall y_1)\ldots(\forall y_n)\mathscr{B} \Rightarrow (\forall y_1)\ldots(\forall y_n)\mathscr{C}]$ where $\mathscr{B}$ and $\mathscr{C}$ are any wfs and $y_1 \ldots, y_n$ are any variables.

Moreover, $K^{\#}$ has modus ponens as its only rule of inference. Show that $K^{\#}$ has the same theorems as K. Thus, at the expense of adding more axioms, the generalization rule can be dispensed with.

2.29 Carry out the proof of the Extension of Propositions 2.4–2.7 above.

2.5 ADDITIONAL METATHEOREMS AND DERIVED RULES

For the sake of smoothness in working with particular theories later, we shall introduce various techniques for constructing proofs. In this section it is assumed that we are dealing with an arbitray theory K.

Often one wants to obtain $\mathscr{B}(t)$ from $(\forall x)\mathscr{B}(x)$, where t is a term free for x in $\mathscr{B}(x)$. This is allowed by the following *derived rule*.

PARTICULARIZATION RULE A4

If t is free for x in $\mathscr{B}(x)$, then $(\forall x)\mathscr{B}(x) \vdash \mathscr{B}(t)$.[†]

Proof

From $(\forall x)\mathscr{B}(x)$ and the instance $(\forall x)\mathscr{B}(x) \Rightarrow \mathscr{B}(t)$ of axiom (A4), we obtain $\mathscr{B}(t)$ by modus ponens.

Since x is free for x in $\mathscr{B}(x)$, a special case of rule A4 is: $(\forall x)\mathscr{B} \vdash \mathscr{B}$.

There is another very useful derived rule, which is essentially the contrapositive of rule A4.

[†]From a strict point of view, $(\forall x)\mathscr{B}(x) \vdash \mathscr{B}(t)$ states a fact about derivability. Rule A4 should be taken to mean that, if $(\forall x)\mathscr{B}(x)$ occurs as a step in a proof, we may write $\mathscr{B}(t)$ as a later step (if t is free for x in $\mathscr{B}(x)$). As in this case, we shall often state a derived rule in the form of the corresponding derivability result that justifies the rule.

EXISTENTIAL RULE E4

Let t be a term that is free for x in a wf $\mathscr{B}(x,t)$, and let $\mathscr{B}(t,t)$ arise from $\mathscr{B}(x,t)$ by replacing all free occurrences of x by t. ($\mathscr{B}(x,t)$ may or may not contain occurrences of t.) Then, $\mathscr{B}(t,t) \vdash (\exists x)\mathscr{B}(x,t)$.

Proof

It suffices to show that $\vdash \mathscr{B}(t,t) \Rightarrow (\exists x)\mathscr{B}(x,t)$. But, by axiom (A4), $\vdash (\forall x)\neg\mathscr{B}(x,t) \Rightarrow \neg\mathscr{B}(t,t)$. Hence, by the tautology $(A \Rightarrow \neg B) \Rightarrow (B \Rightarrow \neg A)$ and MP, $\vdash \mathscr{B}(t,t) \Rightarrow \neg(\forall x)\neg\mathscr{B}(x,t)$, which, in abbreviated form, is $\vdash \mathscr{B}(t,t) \Rightarrow (\exists x)\mathscr{B}(x,t)$.

A special case of rule E4 is $\mathscr{B}(t) \vdash (\exists x)\mathscr{B}(x)$, whenever t is free for x in $\mathscr{B}(x)$. In particular, when t is x itself, $\mathscr{B}(x) \vdash (\exists x)\mathscr{B}(x)$.

Example
$\vdash (\forall x)\mathscr{B} \Rightarrow (\exists x)\mathscr{B}$

1. $(\forall x)\mathscr{B}$	Hyp
2. $\mathscr{B}$	1, rule A4
3. $(\exists x)\mathscr{B}$	2, rule E4
4. $(\forall x)\mathscr{B} \vdash (\exists x)\mathscr{B}$	1–3
5. $\vdash (\forall x)\mathscr{B} \Rightarrow (\exists x)\mathscr{B}$	1–4, Corollary 2.6

The following derived rules are extremely useful.

Negation elimination: $\neg\neg\mathscr{B} \vdash \mathscr{B}$
Negation introduction: $\mathscr{B} \vdash \neg\neg\mathscr{B}$
Conjunction elimination: $\mathscr{B} \wedge \mathscr{C} \vdash \mathscr{B}$
$\mathscr{B} \wedge \mathscr{C} \vdash \mathscr{C}$
$\neg(\mathscr{B} \wedge \mathscr{C}) \vdash \neg\mathscr{B} \vee \neg\mathscr{C}$
Conjunction introduction: $\mathscr{B}, \mathscr{C} \vdash \mathscr{B} \wedge \mathscr{C}$
Disjunction elimination: $\mathscr{B} \vee \mathscr{C}, \neg\mathscr{B} \vdash \mathscr{C}$
$\mathscr{B} \vee \mathscr{C}, \neg\mathscr{C} \vdash \mathscr{B}$
$\neg(\mathscr{B} \vee \mathscr{C}) \vdash \neg\mathscr{B} \wedge \neg\mathscr{C})$
$\mathscr{B} \Rightarrow \mathscr{D}, \mathscr{C} \Rightarrow \mathscr{D}, \mathscr{B} \vee \mathscr{C} \vdash \mathscr{D}$
Disjunction introduction: $\mathscr{B} \vdash \mathscr{B} \vee \mathscr{C}$
$\mathscr{C} \vdash \mathscr{B} \vee \mathscr{C}$
Conditional elimination: $\mathscr{B} \Rightarrow \mathscr{C}, \neg\mathscr{C} \vdash \neg\mathscr{B}$
$\mathscr{B} \Rightarrow \neg\mathscr{C}, \mathscr{C} \vdash \neg\mathscr{B}$
$\neg\mathscr{B} \Rightarrow \mathscr{C}, \neg\mathscr{C} \vdash \mathscr{B}$
$\neg\mathscr{B} \Rightarrow \neg\mathscr{C}, \mathscr{C} \vdash \mathscr{B}$
$\neg(\mathscr{B} \Rightarrow \mathscr{C}) \vdash \mathscr{B}$
$\neg(\mathscr{B} \Rightarrow \mathscr{C}) \vdash \neg\mathscr{C}$
Conditional introduction: $\mathscr{B}, \neg\mathscr{C} \vdash \neg(\mathscr{B} \Rightarrow \mathscr{C})$
Conditional contrapositive: $\mathscr{B} \Rightarrow \mathscr{C} \vdash \neg\mathscr{C} \Rightarrow \neg\mathscr{B}$

$$\neg\mathscr{C} \Rightarrow \neg\mathscr{B} \vdash \mathscr{B} \Rightarrow \mathscr{C}$$

Biconditional elimination: $\mathscr{B} \Leftrightarrow \mathscr{C}, \mathscr{B} \vdash \mathscr{C}$ $\quad \mathscr{B} \Leftrightarrow \mathscr{C}, \neg\mathscr{B} \vdash \neg\mathscr{C}$

$\mathscr{B} \Leftrightarrow \mathscr{C}, \mathscr{C} \vdash \mathscr{B}$ $\quad \mathscr{B} \Leftrightarrow \mathscr{C}, \neg\mathscr{C} \vdash \neg\mathscr{B}$

$\mathscr{B} \Leftrightarrow \mathscr{C} \vdash \mathscr{B} \Rightarrow \mathscr{C}$ $\quad \mathscr{B} \Leftrightarrow \mathscr{C} \vdash \mathscr{C} \Rightarrow \mathscr{B}$

Biconditional introduction: $\mathscr{B} \Rightarrow \mathscr{C}, \mathscr{C} \Rightarrow \mathscr{B} \vdash \mathscr{B} \Leftrightarrow \mathscr{C}$

Biconditional negation: $\mathscr{B} \Leftrightarrow \mathscr{C} \vdash \neg\mathscr{B} \Leftrightarrow \neg\mathscr{C}$

$\neg\mathscr{B} \Leftrightarrow \neg\mathscr{C} \vdash \mathscr{B} \Leftrightarrow \mathscr{C}$

Proof by contradiction: If a proof of $\Gamma, \neg\mathscr{B} \vdash \mathscr{C} \wedge \neg\mathscr{C}$ involves no application of Gen using a variable free in $\mathscr{B}$, then $\Gamma \vdash \mathscr{B}$. (Similarly, one obtains $\Gamma \vdash \neg\mathscr{B}$ from $\Gamma, \mathscr{B} \vdash \mathscr{C} \wedge \neg\mathscr{C}$.)

Exercises

2.30 Justify the derived rules listed above.

2.31 Prove the following.

(a) $\vdash (\forall x)(\forall y)A_1^2(x,y) \Rightarrow (\forall x)A_1^2(x,x)$

(b) $\vdash [(\forall x)\mathscr{B}] \vee [(\forall x)\mathscr{C}] \Rightarrow (\forall x)(\mathscr{B} \vee \mathscr{C})$

(c) $\vdash \neg(\exists x)\mathscr{B} \Rightarrow (\forall x)\neg\mathscr{B}$

(d) $\vdash (\forall x)\mathscr{B} \Rightarrow (\forall x)(\mathscr{B} \vee \mathscr{C})$

(e) $\vdash (\forall x)(\forall y)(A_1^2(x,y) \Rightarrow \neg A_1^2(y,x)) \Rightarrow (\forall x)\neg A_1^2(x,x)$

(f) $\vdash [(\exists x)\mathscr{B} \Rightarrow (\forall x)\mathscr{C}] \Rightarrow (\forall x)(\mathscr{B} \Rightarrow \mathscr{C})$

(g) $\vdash (\forall x)(\mathscr{B} \vee \mathscr{C}) \Rightarrow [(\forall x)\mathscr{B}] \vee (\exists x)\mathscr{C}$

(h) $\vdash (\forall x)(A_1^2(x,x) \Rightarrow (\exists y)A_1^2(x,y))$

(i) $\vdash (\forall x)(\mathscr{B} \Rightarrow \mathscr{C}) \Rightarrow [(\forall x)\neg\mathscr{C} \Rightarrow (\forall x)\neg\mathscr{B}]$

(j) $\vdash (\exists y)[A_1^1(y) \Rightarrow (\forall y)A_1^1(y)]$

(k) $\vdash (\forall x)\mathscr{B} \Rightarrow (\forall x)(\mathscr{B} \vee \mathscr{C})$

(l) $\vdash [(\forall x)(\forall y)(\mathscr{B}(x,y) \Rightarrow \mathscr{B}(y,x)) \wedge (\forall x)(\forall y)(\forall z)(\mathscr{B}(x,y) \wedge \mathscr{B}(y,z) \Rightarrow \mathscr{B}(x,z))] \Rightarrow (\forall x)(\forall y)(\mathscr{B}(x,y) \Rightarrow \mathscr{B}(x,x))$.

2.32 Assume that $\mathscr{B}$ and $\mathscr{C}$ are wfs and that x is not free in $\mathscr{B}$. Prove the following.

(a) $\vdash \mathscr{B} \Rightarrow (\forall x)\mathscr{B}$

(b) $\vdash \mathscr{B} \Rightarrow (\exists x)\mathscr{B}$

(c) $\vdash (\mathscr{B} \Rightarrow (\forall x)\mathscr{C}) \Leftrightarrow (\forall x)(\mathscr{B} \Rightarrow \mathscr{C})$

(d) $\vdash ((\exists x)\mathscr{C} \Rightarrow \mathscr{B}) \Leftrightarrow (\forall x)(\mathscr{C} \Rightarrow \mathscr{B})$

We need a derived rule that will allow us to replace a part $\mathscr{C}$ of a wf $\mathscr{B}$ by a wf that is provably equivalent to $\mathscr{C}$. For this purpose, we first must prove the following auxiliary result.

LEMMA 2.8

For any wfs $\mathscr{B}$ and $\mathscr{C}$, $\vdash (\forall x)(\mathscr{B} \Leftrightarrow \mathscr{C}) \Rightarrow ((\forall x)\mathscr{B} \Leftrightarrow (\forall x)\mathscr{C})$.

Proof

1. $(\forall x)(\mathscr{B} \Leftrightarrow \mathscr{C})$	Hyp
2. $(\forall x)\mathscr{B}$	Hyp
3. $\mathscr{B} \Leftrightarrow \mathscr{C}$	1, rule A4
4. $\mathscr{B}$	2, rule A4
5. $\mathscr{C}$	3, 4, biconditional elimination
6. $(\forall x)\mathscr{C}$	5, Gen
7. $(\forall x)(\mathscr{B} \Leftrightarrow \mathscr{C}), (\forall x)\mathscr{B} \vdash (\forall x)\mathscr{C}$	1–6
8. $(\forall x)(\mathscr{B} \Leftrightarrow \mathscr{C}) \vdash (\forall x)\mathscr{B} \Rightarrow (\forall x)\mathscr{C}$	1–7, Corollary 2.6
9. $(\forall x)(\mathscr{B} \Leftrightarrow \mathscr{C}) \vdash (\forall x)\mathscr{C} \Rightarrow (\forall x)\mathscr{B}$	Proof like that of 8
10. $(\forall x)(\mathscr{B} \Leftrightarrow \mathscr{C}) \vdash (\forall x)\mathscr{B} \Leftrightarrow (\forall x)\mathscr{C}$	8, 9, Bioconditional introduction
11. $\vdash (\forall x)(\mathscr{B} \Leftrightarrow \mathscr{C}) \Rightarrow ((\forall x)\mathscr{B} \Leftrightarrow (\forall x)\mathscr{C})$	1–10, Corollary 2.6

PROPOSITION 2.9

If $\mathscr{C}$ is a subformula of $\mathscr{B}$, $\mathscr{B}'$ is the result of replacing zero or more occurrences of $\mathscr{C}$ in $\mathscr{B}$ by a wf $\mathscr{D}$, and every free variable of $\mathscr{C}$ or $\mathscr{D}$ that is also a bound variable of $\mathscr{B}$ occurs in the list $y_1, \ldots, y_k$, then:

(a) $\vdash [(\forall y_1) \ldots (\forall y_k)(\mathscr{C} \Leftrightarrow \mathscr{D})] \Rightarrow (\mathscr{B} \Leftrightarrow \mathscr{B}')$ (Equivalence theorem)
(b) If $\vdash \mathscr{C} \Leftrightarrow \mathscr{D}$, then $\vdash \mathscr{B} \Leftrightarrow \mathscr{B}'$ (Replacement theorem)
(c) If $\vdash \mathscr{C} \Leftrightarrow \mathscr{D}$ and $\vdash \mathscr{B}$, then $\vdash \mathscr{B}'$

Example
(a) $\vdash (\forall x)(A_1^1(x) \Leftrightarrow (A_2^1(x)) \Rightarrow [(\exists x)A_1^1(x) \Leftrightarrow (\exists x)\, A_2^1(x)]$

Proof

(a) We use induction on the number of connectives and quantifiers in $\mathscr{B}$. Note that, if zero occurrences are replaced, $\mathscr{B}'$ is $\mathscr{B}$ and the wf to be proved is an instance of the tautology $A \Rightarrow (B \Leftrightarrow B)$. Note also that, if $\mathscr{C}$ is identical with $\mathscr{B}$ and this occurrence of $\mathscr{C}$ is replaced by $\mathscr{D}$, the wf to be proved, $[(\forall y_1) \ldots (\forall y_k)(\mathscr{C} \Leftrightarrow \mathscr{D})] \Rightarrow (\mathscr{B} \Leftrightarrow \mathscr{B}')$, is derivable by Exercise 2.27(d). Thus, we may assume that $\mathscr{C}$ is a proper part of $\mathscr{B}$ and that at least one occurrence of $\mathscr{C}$ is replaced. Our inductive hypothesis is that the result holds for all wfs with fewer connectives and quantifiers than $\mathscr{B}$.

Case 1. $\mathscr{B}$ is an atomic wf. Then $\mathscr{C}$ cannot be a proper part of $\mathscr{B}$.

Case 2. $\mathscr{B}$ is $\neg\mathscr{E}$. Let $\mathscr{B}'$ be $\neg\mathscr{E}'$. By inductive hypothesis, $\vdash [(\forall y_1) \ldots (\forall y_k)(\mathscr{C} \Leftrightarrow \mathscr{D})] \Rightarrow (\mathscr{E} \Leftrightarrow \mathscr{E}')$. Hence, by a suitable instance of the tautology $(C \Rightarrow (A \Leftrightarrow B)) \Rightarrow (C \Rightarrow (\neg A \Leftrightarrow \neg B))$ and MP, we obtain $\vdash [(\forall y_1) \ldots (\forall y_k)(\mathscr{C} \Leftrightarrow \mathscr{D})] \Rightarrow (\mathscr{B} \Leftrightarrow \mathscr{B}')$.

Case 3. $\mathscr{B}$ is $\mathscr{E} \Rightarrow \mathscr{F}$. Let $\mathscr{B}'$ be $\mathscr{E}' \Rightarrow \mathscr{F}'$. By inductive hypothesis, $\vdash [(\forall y_1)\dots(\forall y_k)(\mathscr{C} \Leftrightarrow \mathscr{D})] \Rightarrow (\mathscr{E} \Leftrightarrow \mathscr{E}')$ and $\vdash [(\forall y_1)\dots(\forall y_k)(\mathscr{C} \Leftrightarrow \mathscr{D})] \Rightarrow (\mathscr{F} \Leftrightarrow \mathscr{F}')$. Using a suitable instance of the tautology

$$(A \Rightarrow (B \Leftrightarrow C)) \wedge (A \Rightarrow (D \Leftrightarrow E)) \Rightarrow (A \Rightarrow [(B \Rightarrow D) \Leftrightarrow (C \Rightarrow E)])$$

we obtain $\vdash [(\forall y_1)\dots(\forall y_k)(\mathscr{C} \Leftrightarrow \mathscr{D})] \Rightarrow (\mathscr{B} \Leftrightarrow \mathscr{B}')$.

Case 4. $\mathscr{B}$ is $(\forall x)\mathscr{E}$. Let $\mathscr{B}'$ be $(\forall x)\mathscr{E}'$. By inductive hypothesis, $\vdash [(\forall y_1)\dots(\forall y_k)(\mathscr{C} \Leftrightarrow \mathscr{D})] \Rightarrow (\mathscr{E} \Leftrightarrow \mathscr{E}')$. Now, x does not occur free in $(\forall y_1)\dots(\forall y_k)(\mathscr{C} \Leftrightarrow \mathscr{D})$ because, if it did, it would be free in $\mathscr{C}$ or $\mathscr{D}$ and, since it is bound in $\mathscr{B}$, it would be one of $y_1, \dots, y_k$ and it would not be free in $(\forall y_1)\dots(\forall y_k)(\mathscr{C} \Leftrightarrow \mathscr{D})$. Hence, using axiom (A5), we obtain $\vdash (\forall y_1)\dots(\forall y_k)(\mathscr{C} \Leftrightarrow \mathscr{D}) \Rightarrow (\forall x)(\mathscr{E} \Leftrightarrow \mathscr{E}')$. However, by Lemma 2.8, $\vdash (\forall x)(\mathscr{E} \Leftrightarrow \mathscr{E}') \Rightarrow ((\forall x)\mathscr{E} \Leftrightarrow (\forall x)\mathscr{E}')$. Then, by a suitable tautology and MP, $\vdash [(\forall y_1)\dots(\forall y_k)(\mathscr{C} \Leftrightarrow \mathscr{D})] \Rightarrow (\mathscr{B} \Leftrightarrow \mathscr{B}')$.

(b) From $\vdash \mathscr{C} \Leftrightarrow \mathscr{D}$, by several applications of Gen, we obtain $\vdash (\forall y_1)\dots(\forall y_k)(\mathscr{C} \Leftrightarrow \mathscr{D})$. Then, by (a) and MP, $\vdash \mathscr{B} \Leftrightarrow \mathscr{B}'$.

(c) Use part (b) and biconditional elimination.

Exercises

2.33 Prove the following:

(a) $\vdash (\exists x)\neg\mathscr{B} \Leftrightarrow \neg(\forall x)\mathscr{B}$

(b) $\vdash (\forall x)\mathscr{B} \Leftrightarrow \neg(\exists x)\neg\mathscr{B}$

(c) $\vdash (\exists x)(\mathscr{B} \Rightarrow \neg(\mathscr{C} \vee \mathscr{D})) \Rightarrow (\exists x)(\mathscr{B} \Rightarrow \neg\mathscr{C} \wedge \neg\mathscr{D})$

(d) $\vdash (\forall x)(\exists y)(\mathscr{B} \Rightarrow \mathscr{C}) \Leftrightarrow (\forall x)(\exists y)(\neg\mathscr{B} \vee \mathscr{C})$

(e) $\vdash (\forall x)(\mathscr{B} \Rightarrow \neg\mathscr{C}) \Leftrightarrow \neg(\exists x)(\mathscr{B} \wedge \mathscr{C})$

2.34 Show by a counterexample that we cannot omit the quantifiers $(\forall y_1)\dots(\forall y_k)$ in Proposition 2.9(a).

2.35 If $\mathscr{C}$ is obtained from $\mathscr{B}$ by erasing all quantifiers $(\forall x)$ or $(\exists x)$ whose scope does not contain x free, prove that $\vdash \mathscr{B} \Leftrightarrow \mathscr{C}$.

2.36 For each wf $\mathscr{B}$ below, find a wf $\mathscr{C}$ such that $\vdash \mathscr{C} \Leftrightarrow \neg\mathscr{B}$ and negation signs in $\mathscr{C}$ apply only to atomic wfs.

(a) $(\forall x)(\forall y)(\exists z)A_1^3(x, y, z)$

(b) $(\forall\varepsilon)(\varepsilon > 0 \Rightarrow (\exists\delta)(\delta > 0 \wedge (\forall x)(|x - c| < \delta \Rightarrow |f(x) - f(c)| < \varepsilon))$

(c) $(\forall\varepsilon)(\varepsilon > 0 \Rightarrow (\exists n)(\forall m)(m > n \Rightarrow |a_m - b| < \varepsilon))$

2.37 Let $\mathscr{B}$ be a wf that does not contain $\Rightarrow$ and $\Leftrightarrow$. Exchange universal and existential quantifiers and exchange $\wedge$ and $\vee$. The result $\mathscr{B}^*$ is called the *dual* of $\mathscr{B}$.

(a) In any predicate calculus, prove the following.

(i) $\vdash \mathscr{B}$ if and only if $\vdash \neg\mathscr{B}^*$

(ii) $\vdash \mathscr{B} \Rightarrow \mathscr{C}$ if and only if $\vdash \mathscr{C}^* \Rightarrow \mathscr{B}^*$.

(iii) $\vdash \mathscr{B} \Leftrightarrow \mathscr{C}$ if and only if $\vdash \mathscr{B}^* \Leftrightarrow \mathscr{C}^*$.

(iv) $\vdash (\exists x)(\mathscr{B} \vee \mathscr{C}) \Leftrightarrow [((\exists x)\mathscr{B}) \vee (\exists x)\mathscr{C}]$. [*Hint*: Use Exercise 2.27(c).]

(b) Show that the duality results of part (a), (i) (iii), do not hold for arbitrary theories.

2.6 RULE C

It is very common in mathematics to reason in the following way. Assume that we have proved a wf of the form $(\exists x)\mathscr{B}(x)$. Then we say, let b be an object such that $\mathscr{B}(b)$. We continue the proof, finally arriving at a formula that does not involve the arbitrarily chosen element b.

For example, let us say that we wish to show that
$(\exists x)(\mathscr{B}(x) \Rightarrow \mathscr{C}(x)), (\forall x)\mathscr{B}(x) \vdash (\exists x)\mathscr{C}(x)$.

1. $(\exists x)(\mathscr{B}(x) \Rightarrow \mathscr{C}(x))$	Hyp
2. $(\forall x)\mathscr{B}(x)$	Hyp
3. $\mathscr{B}(b) \Rightarrow \mathscr{C}(b)$ for some b	1
4. $\mathscr{B}(b)$	2, rule A4
5. $\mathscr{C}(b)$	3, 4, MP
6. $(\exists x)\mathscr{C}(x)$	5, rule E4

Such a proof seems to be perfectly legitimate on an intuitive basis. In fact, we can achieve the same result without making an arbitrary choice of an element b as in step 3. This can be done as follows:

1. $(\forall x)\mathscr{B}(x)$	Hyp
2. $(\forall x)\neg\mathscr{C}(x)$	Hyp
3. $\mathscr{B}(x)$	1, rule A4
4. $\neg\mathscr{C}(x)$	2, rule A4
5. $\neg(\mathscr{B}(x) \Rightarrow \mathscr{C}(x))$	3, 4, conditional introduction
6. $(\forall x)\neg(\mathscr{B}(x) \Rightarrow \mathscr{C}(x))$	5, Gen
7. $(\forall x)\mathscr{B}(x), (\forall x)\neg\mathscr{C}(x) \vdash$ $(\forall x)\neg(\mathscr{B}(x) \Rightarrow \mathscr{C}(x))$	1- 6
8. $(\forall x)\mathscr{B}(x) \vdash (\forall x)\neg\mathscr{C}(x)$ $\Rightarrow (\forall x)\neg(\mathscr{B}(x) \Rightarrow \mathscr{C}(x))$	1- 7, corollary 2.6
9. $(\forall x)\mathscr{B}(x) \vdash \neg(\forall x)\neg(\mathscr{B}(x)$ $\Rightarrow \mathscr{C}(x)) \Rightarrow \neg(\forall x)\neg\mathscr{C}(x)$	8, contrapositive
10. $(\forall x)\mathscr{B}(x) \vdash (\exists x)(\mathscr{B}(x) \Rightarrow$ $\mathscr{C}(x)) \Rightarrow (\exists x)\mathscr{C}(x)$	Abbreviation of 9
11. $(\exists x)(\mathscr{B}(x) \Rightarrow \mathscr{C}(x)),$ $(\forall x)\mathscr{B}(x) \vdash (\exists x)\mathscr{C}(x)$	10, MP

In general, any wf that can be proved using a finite number of arbitrary choices can also be proved without such acts of choice. We shall call the rule that permits us to go from $(\exists x)\mathscr{B}(x)$ to $\mathscr{B}(b)$, *rule C* ('C' for 'choice'). More precisely, a rule C deduction in a first-order theory K is defined in the

following manner: $\Gamma \vdash_C \mathscr{B}$ if and only if there is a sequence of wfs $\mathscr{D}_1, \cdots, \mathscr{D}_n$ such that $\mathscr{D}_n$ is $\mathscr{B}$ and the following four conditions hold:

1. For each $i < n$, either
   (a) $\mathscr{D}_i$ is an axiom of K, or
   (b) $\mathscr{D}_i$ is in Γ, or
   (c) $\mathscr{D}_i$ follows by MP or Gen from preceding wfs in the sequence, or
   (d) there is a preceding wf $(\exists x)\mathscr{C}(x)$ such that $\mathscr{D}_i$ is $\mathscr{C}(d)$, where d is a new individual constant (rule C).
2. As axioms in condition 1(a), we also can use all logical axioms that involve the new individual constants already introduced in the sequence by applications of rule C.
3. No application of Gen is made using a variable that is free in some $(\exists x)\mathscr{C}(x)$ to which rule C has been previously applied.
4. $\mathscr{B}$ contains none of the new individual constants introduced in the sequence in any application of rule C.

A word should be said about the reason for including condition 3. If an application of rule C to a wf $(\exists x)\mathscr{C}(x)$ yields $\mathscr{C}(d)$, then the object referred to by d may depend on the values of the free variables in $(\exists x)\mathscr{C}(x)$. So that one object may not satisfy $\mathscr{C}(x)$ for *all* values of the free variables in $(\exists x)\mathscr{C}(x)$. For example, without clause 3, we could proceed as follows:

1. $(\forall x)(\exists y)A_1^2(x,y)$	Hyp
2. $(\exists y)A_1^2(x,y)$	1, rule A4
3. $A_1^2(x,d)$	2, rule C
4. $(\forall x)A_1^2(x,d)$	3, Gen
5. $(\exists y)(\forall x)A_1^2(x,y)$	4, rule E4

However, there is an interpretation for which $(\forall x)(\exists y)A_1^2(x,y)$ is true but $(\exists y)(\forall x)A_1^2(x,y)$ is false. Take the domain to be the set of integers and let $A_1^2(x,y)$ mean that $x < y$.

PROPOSITION 2.10

If $\Gamma \vdash_C \mathscr{B}$, then $\Gamma \vdash \mathscr{B}$. Moreover, from the following proof it is easy to verify that, if there is an application of Gen in the new proof of $\mathscr{B}$ from Γ using a certain variable and applied to a wf depending upon a certain wf of Γ, then there was such an application of Gen in the original proof.†

†The first formulation of a version of rule C similar to that given here seems to be due to Rosser (1953).

Proof

Let $(\exists y_1)\mathscr{C}_1(y_1), \ldots, (\exists y_k)\mathscr{C}_k(y_k)$ be the wfs in order of occurrence to which rule C is applied in the proof of $\Gamma \vdash_C \mathscr{B}$, and let $d_1, \ldots, d_k$ be the corresponding new individual constants. Then $\Gamma, \mathscr{C}_1(d_1), \ldots, \mathscr{C}_k(d_k) \vdash \mathscr{B}$. Now, by condition 3 of the definition above, Corollary 2.6 is applicable, yielding $\Gamma, \mathscr{C}_1(d_1), \ldots, \mathscr{C}_{k-1}(d_{k-1}) \vdash \mathscr{C}_k(d_k) \Rightarrow \mathscr{B}$. We replace d_k everywhere by a variable z that does not occur in the proof.
Then

$$\Gamma, \mathscr{C}_1(d_1), \ldots, \mathscr{C}_{k-1}(d_{k-1}) \vdash \mathscr{C}_k(z) \Rightarrow \mathscr{B}$$

and, by Gen,

$$\Gamma, \mathscr{C}_1(d_1), \ldots, \mathscr{C}_{k-1}(d_{k-1}) \vdash (\forall z)(\mathscr{C}_k(z) \Rightarrow \mathscr{B})$$

Hence, by Exercise 2.32(d),

$$\Gamma, \mathscr{C}_1(d_1), \ldots, \mathscr{C}_{k-1}(d_{k-1}) \vdash (\exists y_k)\mathscr{C}_k(y_k) \Rightarrow \mathscr{B}$$

But,

$$\Gamma, \mathscr{C}_1(d_1), \ldots, \mathscr{C}_{k-1}(d_{k-1}) \vdash (\exists y_k)\mathscr{C}_k(y_k)$$

Hence, by MP,

$$\Gamma, \mathscr{C}_1(d_1), \ldots, \mathscr{C}_{k-1}(d_{k-1}) \vdash \mathscr{B}$$

Repeating this argument, we can eliminate $\mathscr{C}_{k-1}(d_{k-1}), \ldots, \mathscr{C}_1(d_1)$ one after the other, finally obtaining $\Gamma \vdash \mathscr{B}$.

Example
$\vdash (\forall x)(\mathscr{B}(x) \Rightarrow \mathscr{C}(x)) \Rightarrow ((\exists x)\mathscr{B}(x) \Rightarrow (\exists x)\mathscr{C}(x))$

1. $(\forall x)(\mathscr{B}x) \Rightarrow \mathscr{C}(x))$	Hyp
2. $(\exists x)\mathscr{B}(x)$	Hyp
3. $\mathscr{B}(d)$	2, rule C
4. $\mathscr{B}(d) \Rightarrow \mathscr{C}(d)$	1, rule A4
5. $\mathscr{C}(d)$	3, 4, MP
6. $(\exists x)\mathscr{C}(x)$	5, rule E4
7. $(\forall x)(\mathscr{B}(x) \Rightarrow \mathscr{C}(x)), (\exists x)\mathscr{B}(x) \vdash_C (\exists x)\mathscr{C}(x)$	$1-6$
8. $(\forall x)(\mathscr{B}(x) \Rightarrow \mathscr{C}(x)), (\exists x)\mathscr{B}(x) \vdash (\exists x)\mathscr{C}(x)$	7, Proposition 2.10
9. $(\forall x)(\mathscr{B}(x) \Rightarrow \mathscr{C}(x)) \vdash (\exists x)\mathscr{B}(x) \Rightarrow (\exists x)\mathscr{C}(x)$	$1-8$, corollary 2.6
10. $\vdash (\forall x)(\mathscr{B}(x) \Rightarrow \mathscr{C}(x)) \Rightarrow ((\exists x)\mathscr{B}(x) \Rightarrow (\exists x)\mathscr{C}(x))$	$1-9$, corollary 2.6

Exercises

Use rule C and Proposition 2.10 to prove Exercises 2.38–2.45.

2.38 $\vdash (\exists x)\mathscr{B}(x) \Rightarrow \mathscr{C}(x)) \Rightarrow ((\forall x)\mathscr{B}(x) \Rightarrow (\exists x)\mathscr{C}(x))$
2.39 $\vdash \neg(\exists y)(\forall x)(A_1^2(x,y) \Leftrightarrow \neg A_1^2(x,x))$

2.40 $\vdash [(\forall x)(A_1^1(x) \Rightarrow A_2^1(x) \vee A_3^1(x)) \wedge \neg(\forall x)(A_1^1(x) \Rightarrow A_2^1(x))] \Rightarrow (\exists x)(A_1^1(x) \wedge A_3^1(x))$

2.41 $\vdash [(\exists x)\mathscr{B}(x)] \wedge [(\forall x)\mathscr{C}(x)] \Rightarrow (\exists x)(\mathscr{B}(x) \wedge \mathscr{C}(x))$

2.42 $\vdash (\exists x)\mathscr{C}(x) \Rightarrow (\exists x)(\mathscr{B}(x) \vee \mathscr{C}(x))$

2.43 $\vdash (\exists x)(\exists y)\mathscr{B}(x,y) \Leftrightarrow (\exists y)(\exists x)\mathscr{B}(x,y)$

2.44 $\vdash (\exists x)(\forall y)\mathscr{B}(x,y) \Rightarrow (\forall y)(\exists x)\mathscr{B}(x,y)$

2.45 $\vdash (\exists x)(\mathscr{B}(x) \wedge \mathscr{C}(x)) \Rightarrow ((\exists x)\mathscr{B}(x)) \wedge (\exists x)\mathscr{C}(x)$

2.46 What is wrong with the following alleged derivations?

(a)

1.	$(\exists x)\mathscr{B}(x)$	Hyp
2.	$\mathscr{B}(d)$	1, rule C
3.	$(\exists x)\mathscr{C}(x)$	Hyp
4.	$\mathscr{C}(d)$	3, rule C
5.	$\mathscr{B}(d) \wedge \mathscr{C}(d)$	2, 4, conjunction introduction
6.	$(\exists x)(\mathscr{B}(x) \wedge \mathscr{C}(x))$	5, rule E4
7.	$(\exists x)\mathscr{B}(x), (\exists x)\mathscr{C}(x) \vdash (\exists x)(\mathscr{B}(x) \wedge \mathscr{C}(x))$	1–6, Proposition 2.10

(b)

1.	$(\exists x)(\mathscr{B}(x) \Rightarrow \mathscr{C}(x))$	Hyp
2.	$(\exists x)\mathscr{B}(x)$	Hyp
3.	$\mathscr{B}(d) \Rightarrow \mathscr{C}(d)$	1, rule C
4.	$\mathscr{B}(d)$	2, rule C
5.	$\mathscr{C}(d)$	3, 4, MP
6.	$(\exists x)\mathscr{C}(x)$	5, rule E4
7.	$(\exists x)(\mathscr{B}(x) \Rightarrow \mathscr{C}(x)), (\exists x)\mathscr{B}(x) \vdash (\exists x)\mathscr{C}(x)$	1–6, Proposition 2.10

2.7 COMPLETENESS THEOREMS

We intend to show that the theorems of a first-order predicate calculus K are precisely the same as the logically valid wfs of K. Half of this result was proved in Proposition 2.2. The other half will follow from a much more general proposition established later. First we must prove a few preliminary lemmas.

If x_i and x_j are distinct, then $\mathscr{B}(x_i)$ and $\mathscr{B}(x_j)$ are said to be *similar* if and only if x_j is free for x_i in $\mathscr{B}(x_i)$ and $\mathscr{B}(x_i)$ has no free occurrences of x_j. It is assumed here that $\mathscr{B}(x_j)$ arises from $\mathscr{B}(x_i)$ by substituting x_j for all free occurrences of x_i. It is easy to see that, if $\mathscr{B}(x_i)$ and $\mathscr{B}(x_j)$ are similar, then x_i is free for x_j in $\mathscr{B}(x_j)$ and $\mathscr{B}(x_j)$ has no free occurrences of x_i. Thus, if $\mathscr{B}(x_i)$ and $\mathscr{B}(x_j)$ are similar, then $\mathscr{B}(x_j)$ and $\mathscr{B}(x_i)$ are similar. Intuitively, $\mathscr{B}(x_i)$ and $\mathscr{B}(x_j)$ are similar if and only if $\mathscr{B}(x_i)$ and $\mathscr{B}(x_j)$ are the same except that $\mathscr{B}(x_i)$ has free occurrences of x_i in exactly those places where $\mathscr{B}(x_j)$ has free occurrences of x_j.

Example

$(\forall x_3)[A_1^2(x_1, x_3) \vee A_1^1(x_1)]$ and $(\forall x_3)[A_1^2(x_2, x_3) \vee A_1^1(x_2)]$ are similar.

LEMMA 2.11

If $\mathscr{B}(x_i)$ and $\mathscr{B}(x_j)$ are similar, then $\vdash (\forall x_i)\mathscr{B}(x_i) \Leftrightarrow (\forall x_j)\mathscr{B}(x_j)$.

Proof

$\vdash (\forall x_i)\mathscr{B}(x_i) \Rightarrow \mathscr{B}(x_j)$ by axiom (A4). Then, by Gen, $\vdash (\forall x_j)((\forall x_i)\mathscr{B}(x_i) \Rightarrow \mathscr{B}(x_j))$, and so, by axiom (A5) and MP, $\vdash (\forall x_i)\mathscr{B}(x_i) \Rightarrow (\forall x_j)\mathscr{B}(x_j)$. Similarly, $\vdash (\forall x_j)\mathscr{B}(x_j) \Rightarrow (\forall x_i)\mathscr{B}(x_i)$. Hence, by biconditional introduction, $\vdash (\forall x_i)\mathscr{B}(x_i) \Leftrightarrow (\forall x_j)\mathscr{B}(x_j)$.

Exercises

2.47 If $\mathscr{B}(x_i)$ and $\mathscr{B}(x_j)$ are similar, prove that $\vdash (\exists x_i)\mathscr{B}(x_i) \Leftrightarrow (\exists x_j)\mathscr{B}(x_j)$.
2.48 *Change of bound variables.* If $\mathscr{B}(x)$ is similar to $\mathscr{B}(y)$, $(\forall x)\mathscr{B}(x)$ is a subformula of $\mathscr{C}$, and $\mathscr{C}'$ is the result of replacing one or more occurrences of $(\forall x)\mathscr{B}(x)$ in $\mathscr{C}$ by $(\forall y)\mathscr{B}(y)$, prove that $\vdash \mathscr{C} \Leftrightarrow \mathscr{C}'$.

LEMMA 2.12

If a closed wf $\neg\mathscr{B}$ of a theory K is not provable in K, and if K′ is the theory obtained from K by adding $\mathscr{B}$ as a new axiom, then K′ is consistent.

Proof

Assume K′ inconsistent. Then, for some wf $\mathscr{C}$, $\vdash_{K'} \mathscr{C}$ and $\vdash_{K'} \neg\mathscr{C}$. Now, $\vdash_{K'} \mathscr{C} \Rightarrow (\neg\mathscr{C} \Rightarrow \neg\mathscr{B})$ by Proposition 2.1. So, by two applications of MP, $\vdash_{K'} \neg\mathscr{B}$. Now, any use of $\mathscr{B}$ as an axiom in a proof in K′ can be regarded as a hypothesis in a proof in K. Hence, $\mathscr{B} \vdash_K \neg\mathscr{B}$. Since $\mathscr{B}$ is closed, we have $\vdash_K \mathscr{B} \Rightarrow \neg\mathscr{B}$ by Corollary 2.7. However, by Proposition 2.1, $\vdash_K (\mathscr{B} \Rightarrow \neg\mathscr{B}) \Rightarrow \neg\mathscr{B}$. Therefore, by MP, $\vdash_K \neg\mathscr{B}$, contradicting our hypothesis.

Exercise

2.49 If a closed wf $\mathscr{B}$ of a theory K is not provable in K, and if K′ is the theory obtained from K by adding $\neg\mathscr{B}$ as a new axiom, then K′ is consistent.

LEMMA 2.13

The set of expressions of a language $\mathscr{L}$ is denumerable. Hence, the same is true of the set of terms, the set of wfs and the set of closed wfs.

Proof

First assign a distinct positive integer $g(u)$ to each symbol u as follows: $g(() = 3,\ g()) = 5,\quad g(,) = 7,\ g(\neg) = 9,\ g(\Rightarrow) = 11,\ g(\forall) = 13,\quad g(x_k) = 13 + 8k,\ g(a_k) = 7 + 8k,\ g(f_k^n) = 1 + 8(2^n 3^k)$, and $g(A_k^n) = 3 + 8(2^n 3^k)$. Then, to an expression $u_0 u_1 \ldots u_r$ associate the number $2^{g(u_0)} 3^{g(u_1)} \ldots p_r^{g(u_r)}$, where p_j is the jth prime number, starting with $p_0 = 2$. (Example: the number of $A_1^1(x_2)$ is $2^{51} 3^3 5^{29} 7^5$.) We can enumerate all expressions in the order of their associated numbers; so, the set of expressions is denumerable.

If we can effectively tell whether any given symbol is a symbol of $\mathscr{L}$, then this enumeration can be effectively carried out, and, in addition, we can effectively decide whether any given number is the number of an expression of $\mathscr{L}$. The same holds true for terms, wfs and closed wfs. If a theory K in the language $\mathscr{L}$ is axiomatic, that is, if we can effectively decide whether any given wf is an axiom of K, then we can effectively enumerate the theorems of K in the following manner. Starting with a list consisting of the first axiom of K in the enumeration just specified, add to the list all the direct consequences of this axiom by MP and by Gen used only once and with x_1 as quantified variable. Add the second axiom to this new list and write all new direct consequences by MP and Gen of the wfs in this augmented list, with Gen used only once and with x_1 and x_2 as quantified variables. If at the kth step we add the kth axiom and apply MP and Gen to the wfs in the new list (with Gen applied only once for each of the variables $x_1, \ldots, x_k$), we eventually obtain in this manner all theorems of K. However, in contradistinction to the case of expressions, terms, wfs and closed wfs, it turns out that there are axiomatic theories K for which we cannot tell in advance whether any given wf of K will eventually appear in the list of theorems.

DEFINITIONS

(i) A theory K is said to be *complete* if, for every closed wf $\mathscr{B}$ of K, either $\vdash_K \mathscr{B}$ or $\vdash_K \neg\mathscr{B}$.

(ii) A theory K′ is said to be an *extension* of a theory K if every theorem of K is a theorem of K′. (We also say in such a case that K is a *subtheory* of K′.)

PROPOSITION 2.14 (LINDENBAUM'S LEMMA)

If K is a consistent theory, then there is a consistent, complete extension of K.

Proof

Let $\mathscr{B}_1, \mathscr{B}_2, \ldots$ be an enumeration of all closed wfs of the language of K, by Lemma 2.13. Define a sequence $J_0, J_1, J_2, \ldots$ of theories in the following way. J_0 is K. Assume J_n is defined, with $n \geqslant 0$. If it is not the case that $\vdash_{J_n} \neg\mathscr{B}_{n+1}$, then let J_{n+1} be obtained from J_n by adding $\mathscr{B}_{n+1}$ as an additional axiom. On the other hand, if $\vdash_{J_n} \neg\mathscr{B}_{n+1}$, let $J_{n+1} = J_n$. Let J be the theory obtained by taking as axioms all the axioms of all the J_is. Clearly, J_{i+1} is an extension of J_i, and J is an extension of all the J_is including $J_0 = K$. To show that J is consistent, it suffices to prove that every J_i is consistent because a proof of a contradiction in J, involving as it does only a finite number of axioms, is also a proof of a contradition in some J_i. We prove the consistency of the J_is, by induction. By hypothesis, $J_0 = K$ is consistent. Assume that J_i is consistent. If $J_{i+1} = J_i$, then J_{i+1} is consistent. If $J_i \neq J_{i+1}$, and therefore, by the definition of J_{i+1}, $\neg\mathscr{B}_{i+1}$ is not provable in J_i, then, by Lemma 2.12, J_{i+1} is also consistent. So, we have proved that all the J_is are consistent and, therefore, that J is consistent. To prove the completeness of J, let $\mathscr{C}$ be any closed wf of K. Then $\mathscr{C} = \mathscr{B}_{j+1}$ for some $j \geqslant 0$. Now, either $\vdash_{J_j} \neg\mathscr{B}_{j+1}$ or $\vdash_{J_{j+1}} \mathscr{B}_{j+1}$, since, if it is not the case that $\vdash_{J_j} \neg\mathscr{B}_{j+1}$, then $\mathscr{B}_{j+1}$ is added as an axiom in J_{j+1}. Therefore, either $\vdash_J \neg\mathscr{B}_{j+1}$ or $\vdash_J \mathscr{B}_{j+1}$. Thus, J is complete.

Note that even if one can effectively determine whether any wf is an axiom of K, it may not be possible to do the same with (or even to enumerate effectively) the axioms of J; that is, J may not be axiomatic even if K is. This is due to the possibility of not being able to determine, at each step, whether or not $\neg\mathscr{B}_{n+1}$ is provable in J_n.

Exercises

2.49 Show that a theory K is complete if and only if, for any closed wfs $\mathscr{B}$ and $\mathscr{C}$ of K, if $\vdash_K \mathscr{B} \vee \mathscr{C}$, then $\vdash_K \mathscr{B}$ or $\vdash_K \mathscr{C}$.

2.50^D Prove that every consistent decidable theory has a consistent, decidable, complete extension.

DEFINITIONS

1. A *closed term* is a term without variables.
2. A theory K is a *scapegoat theory* if, for any wf $\mathscr{B}(x)$ that has x as its only free variable, there is a closed term t such that

$$\vdash_K (\exists x)\, \neg\mathscr{B}(x) \Rightarrow \neg\mathscr{B}(t)$$

LEMMA 2.15

Every consistent theory K has a consistent extension K′ such that K′ is a scapegoat theory and K′ contains denumerably many closed terms.

Proof

Add to the symbols of K a denumerable set $\{b_1, b_2, \ldots\}$ of new individual constants. Call this new theory K_0. Its axioms are those of K plus those logical axioms that involve the symbols of K and the new constants. K_0 is consistent. For, if not, there is a proof in K_0 of a wf $\mathscr{B} \wedge \neg\mathscr{B}$. Replace each b_i appearing in this proof by a variable that does not appear in the proof. This transforms axioms into axioms and preserves the correctness of the applications of the rules of inference. The final wf in the proof is still a contradiction, but now the proof does not involve any of the b_is and therefore is a proof in K. This contradicts the consistency of K. Hence, K_0 is consistent.

By Lemma 2.13, let $F_1(x_{i_1})$, $F_2(x_{i_2}), \ldots, F_k(x_{i_k}), \ldots$ be an enumeration of all wfs of K_0 that have one free variable. Choose a sequence $b_{j_1}, b_{j_2}, \ldots$ of some of the new individual constants such that each b_{j_k} is not contained in any of the wfs $F_1(x_{i_1}), \ldots, F_k(x_{i_k})$ and such that b_{j_k} is different from each of $b_{j_1}, \ldots, b_{j_{k-1}}$. Consider the wf

$$(S_k) \quad (\exists x_{i_k})\neg F_k(x_{i_k}) \Rightarrow \neg F_k(b_{j_k})$$

Let K_n be the theory obtained by adding $(S_1), \ldots, (S_n)$ to the axioms of K_0, and let K_∞ be the theory obtained by adding all the (S_i)s as axioms to K_0. Any proof in K_∞ contains only a finite number of the (S_i)s and, therefore, will also be a proof in some K_n. Hence, if all the K_ns are consistent, so is K_∞. To demonstrate that all the K_ns are consistent, proceed by induction. We know that K_0 is consistent. Assume that K_{n-1} is consistent but that K_n is inconsistent ($n \geqslant 1$). Then, as we know, any wf is provable in K_n (by the tautology $\neg A \Rightarrow (A \Rightarrow B)$, Proposition 2.1 and MP). In particular, $\vdash_{K_n} \neg(S_n)$. Hence, $(S_n) \vdash_{K_{n-1}} \neg(S_n)$. Since (S_n) is closed, we have, by Corollary 2.7, $\vdash_{K_{n-1}} (S_n) \Rightarrow \neg(S_n)$. But, by the tautology $(A \Rightarrow \neg A) \Rightarrow \neg A$, Proposition 2.1 and MP, we then have $\vdash_{K_{n-1}} \neg(S_n)$; that is, $\vdash_{K_{n-1}} \neg[(\exists x_{i_n})\neg F_n(x_{i_n}) \Rightarrow \neg F_n(b_{j_n})]$. Now, by conditional elimination, we obtain $\vdash_{K_{n-1}} (\exists x_{i_n})\neg F_n(x_{i_n})$ and $\vdash_{K_{n-1}} \neg\neg F_n(b_{j_n})$, and then, by negation elimination, $\vdash K_1 F_n(b_{j_n})$. From the latter and the fact that b_{j_n} does not occur in $(S_0), \ldots, (S_{n-1})$, we conclude $\vdash_{K_{n-1}} F_n(x_r)$, where x_r is a variable that does not occur in the proof of $F_n(b_{j_n})$. (Simply replace in the proof all occurrences of b_{j_n} by x_r.) By Gen, $\vdash_{K_{n-1}} (\forall x_r) F_n(x_r)$, and then, by Lemma 2.11 and biconditional elimination, $\vdash_{K_{n-1}} (\forall x_{i_n}) F_n(x_{i_n})$. (We use the fact that $F_n(x_r)$ and $F_n(x_{i_n})$ are similar.) But we already have $\vdash_{K_{n-1}} (\exists x_{i_n})\neg F_n(x_{i_n})$, which is an abbreviation of $\vdash_{K_{n-1}} \neg(\forall x_{i_n})\neg\neg F_n(x_{i_n})$, whence, by the replacement theorem, $\vdash_{K_{n-1}} \neg(\forall x_{i_n}) F_n(x_{i_n})$, contradicting the hypothesis that K_{n-1} is consistent. Hence, K_n must also be consistent. Thus K_∞ is consistent, it is an extension of K, and it is clearly a scapegoat theory.

LEMMA 2.16

Let J be a consistent, complete scapegoat theory. Then J has a model M whose domain is the set D of closed terms of J.

Proof

For any individual constant a_i of J, let $(a_i)^{\mathrm{M}} = a_i$. For any function letter f_k^n of J and for any closed terms $t_1, \ldots t_n$ of J, let $(f_k^n)^{\mathrm{M}}(t_1, \ldots, t_n) = f_k^n(t_1, \ldots, t_n)$. (Notice that $f_k^n(t_1, \ldots, t_n)$ is a closed term. Hence, $(f_k^n)^{\mathrm{M}}$ is an n-ary operation on D.) For any predicate letter A_k^n of J, let $(A_k^n)^{\mathrm{M}}$ consist of all n-tuples $\langle t_1, \ldots, t_n \rangle$ of closed terms $t_1, \ldots, t_n$ of J such that $\vdash_{\mathrm{J}} A_k^n(t_1, \ldots, t_n)$. It now suffices to show that, for any closed wf $\mathscr{C}$ of J:

$$(\square) \qquad \models_{\mathrm{M}} \mathscr{C} \quad \text{if and only if} \quad \vdash_{\mathrm{J}} \mathscr{C}$$

(If this is established and $\mathscr{B}$ is any axiom of J, let $\mathscr{C}$ be the closure of $\mathscr{B}$. By Gen, $\vdash_{\mathrm{J}} \mathscr{C}$. By $(\square)$, $\models_{\mathrm{M}} \mathscr{C}$. By (VI) on page 61, $\models_{\mathrm{M}} \mathscr{B}$. Hence, M would be a model of J.) The proof of $(\square)$ is by induction on the number r of connectives and quantifiers in $\mathscr{C}$. Assume that $(\square)$ holds for all closed wfs with fewer than r connectives and quantifiers.

Case 1. $\mathscr{C}$ is a closed atomic wf $A_k^n(t_1, \ldots, t_n)$. Then $(\square)$ is a direct consequence of the definition of $(A_k^n)^{\mathrm{M}}$.

Case 2. $\mathscr{C}$ is $\neg\mathscr{D}$. If $\mathscr{C}$ is true for M, then $\mathscr{D}$ is false for M and so, by inductive hypothesis, not-$\vdash_{\mathrm{J}} \mathscr{D}$. Since J is complete and $\mathscr{D}$ is closed, $\vdash_{\mathrm{J}} \neg\mathscr{D}$ – that is, $\vdash_{\mathrm{J}} \mathscr{C}$. Conversely, if $\mathscr{C}$ is not true for M, then $\mathscr{D}$ is true for M. Hence, $\vdash_{\mathrm{J}} \mathscr{D}$. Since J is consistent, not-$\vdash_{\mathrm{J}} \neg\mathscr{D}$, that is, not-$\vdash_{\mathrm{J}} \mathscr{C}$.

Case 3. $\mathscr{C}$ is $\mathscr{D} \Rightarrow \mathscr{E}$. Since $\mathscr{C}$ is closed, so are $\mathscr{D}$ and $\mathscr{E}$. If $\mathscr{C}$ is false for M, then $\mathscr{D}$ is true and $\mathscr{E}$ is false. Hence, by inductive hypothesis, $\vdash_{\mathrm{J}} \mathscr{D}$ and not-$\vdash_{\mathrm{J}} \mathscr{E}$. By the completeness of J, $\vdash_{\mathrm{J}} \neg\mathscr{E}$. Therefore, by an instance of the tautology $D \Rightarrow (\neg E \Rightarrow \neg(D \Rightarrow E))$ and two applications of MP, $\vdash_{\mathrm{J}} \neg(\mathscr{D} \Rightarrow \mathscr{E})$, that is, $\vdash_{\mathrm{J}} \neg\mathscr{C}$, and so, by the consistency of J, not-$\vdash_{\mathrm{J}} \mathscr{C}$. Conversely, if not-$\vdash_{\mathrm{J}} \mathscr{C}$, then, by the completeness of J, $\vdash_{\mathrm{J}} \neg\mathscr{C}$, that is, $\vdash_{\mathrm{J}} \neg(\mathscr{D} \Rightarrow \mathscr{E})$. By conditional elimination, $\vdash_{\mathrm{J}} \mathscr{D}$ and $\vdash_{\mathrm{J}} \neg\mathscr{E}$. Hence, by $(\square)$ for $\mathscr{D}$, $\mathscr{D}$ is true for M. By the consistency of J, not-$\vdash_{\mathrm{J}} \mathscr{E}$ and, therefore, by $(\square)$ for $\mathscr{E}$, $\mathscr{E}$ is false for M. Thus, since $\mathscr{D}$ is true for M and $\mathscr{E}$ is false for M, $\mathscr{C}$ is false for M.

Case 4. $\mathscr{C}$ is $(\forall x_m)\mathscr{D}$.

Case 4a. $\mathscr{D}$ is a closed wf. By inductive hypothesis, $\models_{\mathrm{M}} \mathscr{D}$ if and only if $\vdash_{\mathrm{J}} \mathscr{D}$. By Exercise 2.32(a), $\vdash_{\mathrm{J}} \mathscr{D} \Leftrightarrow (\forall x_m)\mathscr{D}$. So, $\vdash_{\mathrm{J}} \mathscr{D}$ if and only if $\vdash_{\mathrm{J}} (\forall x_m)\mathscr{D}$, by biconditional elimination. Moreover, $\models_{\mathrm{M}} \mathscr{D}$ if and only if $\models_{\mathrm{M}} (\forall x_m)\mathscr{D}$ by property (VI) on page 61. Hence, $\models_{\mathrm{M}} \mathscr{C}$ if and only if $\vdash_{\mathrm{J}} \mathscr{C}$.

Case 4b. $\mathscr{D}$ is not a closed wf. Since $\mathscr{C}$ is closed, $\mathscr{D}$ has x_m as its only free variable, say $\mathscr{D}$ is $F(x_m)$. Then $\mathscr{C}$ is $(\forall x_m)F(x_m)$.

(i) Assume $\models_M \mathscr{C}$ and not-$\vdash_J \mathscr{C}$. By the completeness of J, $\vdash_J \neg\mathscr{C}$, that is, $\vdash_J \neg(\forall x_m)F(x_m)$. Then, by Exercise 2.33(a) and biconditional elimination, $\vdash_J (\exists x_m)\neg F(x_m)$. Since J is a scapegoat theory, $\vdash_J \neg F(t)$ for some closed term t of J. But $\models_M \mathscr{C}$, that is, $\models_M (\forall x_m)F(x_m)$. Since $(\forall x_m)F(x_m) \Rightarrow F(t)$ is true for M by property (X) on page 63, $\models_M F(t)$. Hence, by $(\square)$ for $F(t)$, $\vdash_J F(t)$. This contradicts the consistency of J. Thus, if $\models_M \mathscr{C}$, then, $\vdash_J \mathscr{C}$.

(ii) Assume $\vdash_J \mathscr{C}$ and not-$\models_M \mathscr{C}$. Thus,

$$(\#) \quad \vdash_J (\forall x_m)F(x_m)(\#\#) \quad \text{not-} \models_M (\forall x_m)F(x_m).$$

By $(\#\#)$, some sequence of elements of the domain D does not satisfy $(\forall x_m)F(x_m)$. Hence, some sequence s does not satisfy $F(x_m)$. Let t be the ith component of s. Notice that $s^*(u) = u$ for all closed terms u of J (by the definition of $(a_i)^M$ and $(f_k^n)^M$). Observe also that $F(t)$ has fewer connectives and quantifiers than $\mathscr{C}$ and, therefore, the inductive hypothesis applies to $F(t)$, that is, $(\square)$ holds for $F(t)$. Hence, by Lemma 2(a) on page 63, s does not satisfy $F(t)$. So, $F(t)$ is false for M. But, by $(\#)$ and rule A4, $\vdash_J F(t)$, and so, by $(\square)$ for $F(t)$, $\models_M F(t)$. This contradiction shows that, if $\vdash_J \mathscr{C}$, then $\models_M \mathscr{C}$.

Now we can prove the fundamental theorem of quantification theory. By a *denumerable model* we mean a model in which the domain is denumerable.

PROPOSITION 2.17[†]

Every consistent theory K has a denumerable model.

Proof

By Lemma 2.15, K has a consistent extension K′ such that K′ is a scapegoat theory and has denumerably many closed terms. By Lindenbaum's lemma, K′ has a consistent, complete extension J that has the same symbols at K′. Hence, J is also a scapegoat theory. By Lemma 2.16, J has a model M whose domain is the denumerable set of closed terms of J. Since J is an extension of K, M is a denumerable model of K.

[†]The proof given here is essentially due to Henkin (1949), as simplified by Hasenjaeger (1953). The result was originally proved by Gödel (1930). Other proofs have been published by Rasiowa and Sikorski (1951; 1952) and Beth (1951), using (Boolean) algebraic and topological methods, respectively. Still other proofs may be found in Hintikka (1955a, b) and in Beth (1959).

COROLLARY 2.18

Any logically valid wf $\mathscr{B}$ of a theory K is a theorem of K.

Proof

We need consider only closed wfs $\mathscr{B}$, since a wf $\mathscr{D}$ is logically valid if and only if its closure is logically valid, and $\mathscr{D}$ is provable in K if and only if its closure is provable in K. So, let $\mathscr{B}$ be a logically valid closed wf of K. Assume that not-$\vdash_K \mathscr{B}$. By Lemma 2.12, if we add $\neg\mathscr{B}$ as a new axiom to K, the new theory K′ is consistent. Hence, by Proposition 2.17, K′ has a model M. Since $\neg\mathscr{B}$ is an axiom of K′, $\neg\mathscr{B}$ is true for M. But, since $\mathscr{B}$ is logically valid, $\mathscr{B}$ is true for M. Hence, $\mathscr{B}$ is both true and false for M, which is impossible (by (II) on page 61). Thus, $\mathscr{B}$ must be a theorem of K.

COROLLARY 2.19. (GÖDEL'S COMPLETENESS THEOREM, 1930)

In any predicate calculus, the theorems are precisely the logically valid wfs.

Proof

This follows from Proposition 2.2 and Corollary 2.18. (Gödel's original proof runs along quite different lines. For other proofs, see Beth (1951), Dreben (1952), Hintikka (1955a, b) and Rasiowa and Sikorski (1951; 1952).)

COROLLARY 2.20

Let K be any theory.

(a) A wf $\mathscr{B}$ is true in every denumerable model of K if and only if $\vdash_K \mathscr{B}$.
(b) If, in every model of K, every sequence that satisfies all wfs in a set Γ of wfs also satisfies a wf $\mathscr{B}$, then $\Gamma \vdash_K \mathscr{B}$.
(c) If a wf $\mathscr{B}$ of K is a logical consequence of a set Γ of wfs of K, then $\Gamma \vdash_K \mathscr{B}$.
(d) If a wf $\mathscr{B}$ of K is a logical consequence of a wf $\mathscr{C}$ of K, then $\mathscr{C} \vdash_K \mathscr{B}$.

Proof

(a) We may assume $\mathscr{B}$ is closed. If not-$\vdash_K \mathscr{B}$, then the theory $K' = K + \{\neg\mathscr{B}\}$ is consistent.† Hence, by Proposition 2.17, K′ has a denumerable model M. However, $\neg\mathscr{B}$, being an axiom of K′, is true for

†If K is a theory and Δ is a set of wfs of K, then K + Δ denotes the theory obtained from K by adding the wfs of Δ as axioms.

M. By hypothesis, since M is a denumerable model of K, $\mathscr{B}$ is true for M. Therefore, $\mathscr{B}$ is true and false for M, which is impossible.

(b) Consider the theory $K + \Gamma$. By the hypothesis, $\mathscr{B}$ is true for every model of this theory. Hence, by (a), $\vdash_{K+\Gamma} \mathscr{B}$. So, $\Gamma \vdash_K \mathscr{B}$.

Part (c) is a consequence of (b), and part (d) is a special case of (c).

Corollaries 2.18–2.20 show that the 'syntactical' approach to quantification theory by means of first-order theories is equivalent to the 'semantical' approach through the notions of interpretations, models, logical validity, and so on. For the propositional calculus, Corollary 1.15 demonstrated the analogous equivalence between the semantical notion (tautology) and the syntactical notion (theorem of L). Notice also that, in the propositional calculus, the completeness of the system L (see Proposition 1.14) led to a solution of the decision problem. However, for first-order theories, we cannot obtain a decision procedure for logical validity or, equivalently, for provability in first-order predicate calculi. We shall prove this and related results in Section 3.6.

COROLLARY 2.21. (SKOLEM–LÖWENHEIM THEOREM, 1920, 1915)

Any theory that has a model has a denumerable model.

Proof

If K has a model, then K is consistent, since no wf can be both true and false for the same model M. Hence, by Proposition 2.17, K has a denumerable model.

The following stronger consequence of Proposition 2.17 is derivable.

COROLLARY 2.22[A]

For any cardinal number $\mathfrak{m} \geqslant \aleph_0$, any consistent theory K has a model of cardinality $\mathfrak{m}$.

Proof

By Proposition 2.17, we know that K has a denumerable model. Therefore, it suffices to prove the following lemma.

LEMMA

If $\mathfrak{m}$ and $\mathfrak{n}$ are two cardinal numbers such that $\mathfrak{m} \leqslant \mathfrak{n}$ and if K has a model of cardinality $\mathfrak{m}$, then K has a model of cardinality $\mathfrak{n}$.

Proof

Let M be a model of K with domain D of cardinality $\mathfrak{m}$. Let D' be a set of cardinality $\mathfrak{n}$ that contains D. Extend the model M to an interpretation M′ that has D' as domain in the following way. Let c be a fixed element of D. We stipulate that the elements of $D' - D$ behave like c. For example, if B_j^n is the interpretation in M of the predicate letter A_j^n and $(B_j^n)'$ is the new interpretation in M′, then for any $d_1, \ldots, d_n$ in D′, $(B_j^n)'$ holds for $(d_1, \ldots, d_n)$ if and only if B_j^n holds for $(u_1, \ldots, u_n)$, where $u_i = d_i$ if $d_i \in$ D and $u_i = c$ if $d_i \in D' - D$. The interpretation of the function letters is extended in an analogous way, and the individual constants have the same interpretations as in M. It is an easy exercise to show, by induction on the number of connectives and quantifiers in a wf $\mathscr{B}$, that $\mathscr{B}$ is true for M′ if and only if it is true for M. Hence, M′ is a model of K of cardinality $\mathfrak{n}$.

Exercises

2.51 For any theory K, if $\Gamma \vdash_K \mathscr{B}$ and each wf in Γ is true for a model M of K, show that $\mathscr{B}$ is true for M.

2.52 If a wf $\mathscr{B}$ without quantifiers is provable in a predicate calculus, prove that $\mathscr{B}$ is an instance of a tautology and, hence, by Proposition 2.1, has a proof without quantifiers using only axioms (A1)–(A3) and MP. [*Hint*: if $\mathscr{B}$ were not a tautology, one could construct an interpretation, having the set of terms that occur in $\mathscr{B}$ as its domain, for which $\mathscr{B}$ is not true, contradicting Proposition 2.2.] Note that this implies the consistency of the predicate calculus and also provides a decision procedure for the provability of wfs without quantifiers.

2.53 Show that $\vdash_K \mathscr{B}$ if and only if there is a wf $\mathscr{C}$ that is the closure of the conjunction of some axioms of K such that $\mathscr{C} \Rightarrow \mathscr{B}$ is logically valid.

2.54 *Compactness.* If all finite subsets of the set of axioms of a theory K have models, prove that K has a model.

2.55 (a) For any wf $\mathscr{B}$, prove that there is only a finite number of interpretations of $\mathscr{B}$ on a given domain of finite cardinality k.

(b) For any wf $\mathscr{B}$, prove that there is an effective way of determining whether $\mathscr{B}$ is true for all interpretations with domain of some fixed cardinality k.

(c) Let a wf $\mathscr{B}$ be called *k-valid* if it is true for all interpretations that have a domain of k elements. Call $\mathscr{B}$ *precisely k-valid* if it is k-valid but not $(k+1)$-valid. Show that $(k+1)$-validity implies k-validity and give an example of a wf that is precisely k-valid. (See Hilbert and Bernays (1934, § 4–5) and Wajsberg (1933).)

2.56 Show that the following wf is true for all finite domains but is false for some infinite domain.

$$(\forall x)(\forall y)(\forall z)[A_1^2(x,x) \wedge (A_1^2(x,y) \wedge A_1^2(y,z) \Rightarrow A_1^2(x,z)) \wedge (A_1^2(x,y) \vee A_1^2(y,x))] \Rightarrow (\exists y)(\forall x)A_1^2(y,x)$$

2.57 Prove that there is no theory K whose models are exactly the interpretations with finite domains.

2.58 Let $\mathscr{B}$ be any wf that contains no quantifiers, function letters, or individual constants.

(a) Show that a closed *prenex* wf $(\forall x_1)\ldots(\forall x_n)(\exists y_1)\ldots(\exists y_m)\mathscr{B}$, with $m \geqslant 0$ and $n \geqslant 1$, is logically valid if and only if it is true for every interpretation with a domain of n objects.

(b) Prove that a closed prenex wf $(\exists y_1)\ldots(\exists y_m)\mathscr{B}$ is logically valid if and only if it is true for all interpretations with a domain of one element.

(c) Show that there is an effective procedure to determine the logical validity of all wfs of the forms given in (a) and (b).

2.59 Let K_1 and K_2 be theories in the same language $\mathscr{L}$. Assume that any interpretation M of $\mathscr{L}$ is a model of K_1 if and only if M is not a model of K_2. Prove that K_1 and K_2 are finitely axiomatizable, that is, there are finite sets of sentences Γ and Δ such that, for any sentence $\mathscr{B}$, $\vdash_{K_1} \mathscr{B}$ if and only if $\Gamma \vdash \mathscr{B}$, and $\vdash_{K_2} \mathscr{B}$ if and only if $\Delta \vdash \mathscr{B}$.†

2.60 A set Γ of sentences is called an *independent axiomatization* of a theory K if (a) all sentences in Γ are theorems of K, (b) $\Gamma \vdash \mathscr{B}$ for every theorem $\mathscr{B}$ of K, and (c) for every sentence $\mathscr{C}$ of Γ, it is not the case that $\Gamma - \{\mathscr{C}\} \vdash \mathscr{C}$ †. Prove that every theory K has an independent axiomatization.

2.61[A] If, for some cardinal $\mathfrak{m} \geqslant \aleph_0$, a wf $\mathscr{B}$ is true for every interpretation of cardinality $\mathfrak{m}$, prove that $\mathscr{B}$ is logically valid.

2.62[A] If a wf $\mathscr{B}$ is true for all interpretations of cardinality $\mathfrak{m}$ prove that $\mathscr{B}$ is true for all interpretations of cardinality less than or equal to $\mathfrak{m}$.

2.63 (a) Prove that a theory K is a scapegoat theory if and only if, for any wf $\mathscr{B}(x)$ with x as its only free variable, there is a closed term t such that $\vdash_K (\exists x)\mathscr{B}(x) \Rightarrow \mathscr{B}(t)$.

(b) Prove that a theory K is a scapegoat theory if and only if, for any wf $\mathscr{B}(x)$ with x as its only free variable such that $\vdash_K (\exists x)\mathscr{B}(x)$, there is a closed term t such that $\vdash_K \mathscr{B}(t)$.

(c) Prove that no predicate calculus is a scapegoat theory.

2.8 FIRST-ORDER THEORIES WITH EQUALITY

Let K be a theory that has as one of its predicate letters A_1^2. Let us write $t = s$ as an abbreviation for $A_1^2(t,s)$, and $t \neq s$ as an abbreviation for $\neg A_1^2(t,s)$.

†Here, an expression $\Gamma \vdash \mathscr{B}$, without any subscript attached to $\vdash$, means that $\mathscr{B}$ is derivable from Γ using only logical axioms, that is within the predicate calculus.

Then K is called a *first-order theory with equality* (or simply a *theory with equality*) if the following are theorems of K:

(A6) $(\forall x_1)x_1 = x_1$ (reflexivity of equality)
(A7) $x = y \Rightarrow (\mathscr{B}(x,x) \Rightarrow \mathscr{B}(x,y))$ (substitutivity of equality)

where x and y are any variables, $\mathscr{B}(x,x)$ is any wf, and $\mathscr{B}(x,y)$ arises from $\mathscr{B}(x,x)$ by replacing some, but not necessarily all, free occurrences of x by y, with the proviso that y is free for x in $\mathscr{B}(x,x)$. Thus, $\mathscr{B}(x,y)$ may or may not contain free occurrences of x.

The numbering (A6) and (A7) is a continuation of the numbering of the logical axioms.

PROPOSITION 2.23

In any theory with equality,

(a) $\vdash t = t$ for any term t;
(b) $\vdash t = s \Rightarrow s = t$ for any terms t and s;
(c) $\vdash t = s \Rightarrow (s = r \Rightarrow t = r)$ for any terms t, s and r.

Proof

(a) By (A6), $\vdash (\forall x_1)\, x_1 = x_1$. Hence, by rule A4, $\vdash t = t$.
(b) Let x and y be variables not occurring in t or s. Letting $\mathscr{B}(x,x)$ be $x = x$ and $\mathscr{B}(x,y)$ be $y = x$ in schema (A7), $\vdash x = y \Rightarrow (x = x \Rightarrow y = x)$. But, by (a), $\vdash x = x$. So, by an instance of the tautology $(A \Rightarrow (B \Rightarrow C)) \Rightarrow (B \Rightarrow (A \Rightarrow C))$ and two applications of MP, we have $\vdash x = y \Rightarrow y = x$. Two applications of Gen yield $\vdash (\forall x)(\forall y)(x = y \Rightarrow y = x)$, and then two applications of rule A4 give $\vdash t = s \Rightarrow s = t$.
(c) Let x, y and z be three variables not occurring in t, s, or r. Letting $\mathscr{B}(y,y)$ be $y = z$ and $\mathscr{B}(y,x)$ be $x = z$ in (A7), with x and y interchanged, we obtain $\vdash y = x \Rightarrow (y = z \Rightarrow x = z)$. But, by (b), $\vdash x = y \Rightarrow y = x$. Hence, using an instance of the tautology $(A \Rightarrow B) \Rightarrow ((B \Rightarrow C) \Rightarrow (A \Rightarrow C))$ and two applications of MP, we obtain $\vdash x = y \Rightarrow (y = z \Rightarrow x = z)$. By three applications of Gen, $\vdash (\forall x)(\forall y)(\forall z)(x = y \Rightarrow (y = z \Rightarrow x = z))$, and then, by three uses of rule A4, $\vdash t = s \Rightarrow (s = r \Rightarrow t = r)$.

Exercises

2.64 Show that (A6) and (A7) are true for any interpretation M in which $(A_1^2)^{\mathrm{M}}$ is the identity relation on the domain of the interpretation.
2.65 Prove the following in any theory with equality.

(a) $\vdash (\forall x)(\mathscr{B}(x) \Leftrightarrow (\exists y)(x = y \wedge \mathscr{B}(y)))$ if y does not occur in $\mathscr{B}(x)$
(b) $\vdash (\forall x)(\mathscr{B}(x) \Leftrightarrow (\forall y)(x = y \Rightarrow \mathscr{B}(y)))$ if y does not occur in $\mathscr{B}(x)$
(c) $\vdash (\forall x)(\exists y)\ x = y$
(d) $\vdash x = y \Rightarrow f(x) = f(y)$, where f is any function letter of one argument
(e) $\vdash \mathscr{B}(x) \wedge x = y \Rightarrow \mathscr{B}(y)$, if y is free for x in $\mathscr{B}(x)$
(f) $\vdash \mathscr{B}(x) \wedge \neg\mathscr{B}(y) \Rightarrow x \neq y$, if y is free for x in $\mathscr{B}(x)$

We can reduce schema (A7) to a few simpler cases.

PROPOSITION 2.24

Let K be a theory for which (A6) holds and (A7) holds for all atomic wfs $\mathscr{B}(x,x)$ in which there are no individual constants. Then K is a theory with equality, that is, (A7) holds for all wfs $\mathscr{B}(x,x)$.

Proof

We must prove (A7) for all wfs $\mathscr{B}(x,x)$. It holds for atomic wfs by assumption. Note that we have the results of Proposition 2.23, since its proof used (A7) only with atomic wfs without individual constants. Note also that we have (A7) for all atomic wfs $\mathscr{B}(x,x)$. For if $\mathscr{B}(x,x)$ contains individual constants, we can replace those individual constants by new variables, obtaining a wf $\mathscr{B}^*(x,x)$ without individual constants. By hypothesis, the corresponding instance of (A7) with $\mathscr{B}^*(x,x)$ is a theorem; we can then apply Gen with respect to the new variables, and finally apply rule A4 one or more times to obtain (A7) with respect to $\mathscr{B}(x,x)$.

Proceeding by induction on the number n of connectives and quantifiers in $\mathscr{B}(x,x)$, we assume that (A7) holds for all $k < n$.

Case 1. $\mathscr{B}(x,x)$ is $\neg\mathscr{C}(x,x)$. By inductive hypothesis, we have $\vdash y = x \Rightarrow (\mathscr{C}(x,y) \Rightarrow \mathscr{C}(x,x))$, since $\mathscr{C}(x,x)$ arises from $\mathscr{C}(x,y)$ by replacing some occurrences of y by x. Hence, by Proposition 2.23(b), instances of the tautologies $(A \Rightarrow B) \Rightarrow (\neg\mathscr{B} \Rightarrow \neg A)$ and $(A \Rightarrow B) \Rightarrow ((B \Rightarrow C) \Rightarrow (A \Rightarrow C))$ and MP, we obtain $\vdash x = y \Rightarrow (\mathscr{B}(x,x) \Rightarrow \mathscr{B}(x,y))$.

Case 2. $\mathscr{B}(x,x)$ is $\mathscr{C}(x,x) \Rightarrow \mathscr{D}(x,x)$. By inductive hypothesis and Proposition 2.23(b), $\vdash x = y \Rightarrow (\mathscr{C}(x,y) \Rightarrow \mathscr{C}(x,x))$ and $\vdash x = y \Rightarrow (\mathscr{D}(x,x) \Rightarrow \mathscr{D}(x,y))$. Hence, by the tautology $(A \Rightarrow (C_1 \Rightarrow C)) \Rightarrow [(A \Rightarrow (D \Rightarrow D_1)) \Rightarrow (A \Rightarrow ((C \Rightarrow D) \Rightarrow (C_1 \Rightarrow D_1)))]$, we have $\vdash x = y \Rightarrow (\mathscr{B}(x,x) \Rightarrow \mathscr{B}(x,y))$.

Case 3. $\mathscr{B}(x,x)$ is $(\forall z)\mathscr{C}(x,x,z)$. By inductive hypothesis, $\vdash x = y \Rightarrow (\mathscr{C}(x,x,z) \Rightarrow \mathscr{C}(x,y,z))$. Now, by Gen and axiom (A5), $\vdash x = y \Rightarrow (\forall z)(\mathscr{C}(x,x,z) \Rightarrow \mathscr{C}(x,y,z))$. By Exercise 2.27(a), $\vdash (\forall z)\,(\mathscr{C}(x,x,z) \Rightarrow \mathscr{C}(x,y,z)) \Rightarrow [(\forall z)\mathscr{C}(x,x,z) \Rightarrow (\forall z)\mathscr{C}(x,y,z)]$, and so, by the tautology $(A \Rightarrow B) \Rightarrow ((B \Rightarrow C) \Rightarrow (A \Rightarrow C))$, $\vdash x = y \Rightarrow (\mathscr{B}(x,x) \Rightarrow \mathscr{B}(x,y))$.

The instances of (A7) can be still further reduced.

PROPOSITION 2.25

Let K be a theory in which (A6) holds and the following are true.

(a) Schema (A7) holds for all atomic wfs $\mathscr{B}(x,x)$ such that no function letters or individual constants occur in $\mathscr{B}(x,x)$ and $\mathscr{B}(x,y)$ comes from $\mathscr{B}(x,x)$ by replacing exactly one occurrence of x by y.

(b) $\vdash x = y \Rightarrow f_j^n(z_1,\ldots,z_n) = f_j^n(w_1,\ldots,w_n)$, where f_j^n is any function letter of K, $z_1,\ldots,z_n$ are variables, and $f_j^n(w_1,\ldots,w_n)$ arises from $f_j^n(z_1,\ldots,z_n)$ by replacing exactly one occurrence of x by y.

Then K is a theory with equality.

Proof

By repeated application, our assumptions can be extended to replacements of more than one occurrence of x by y. Also, Proposition 2.23 is still derivable. By Proposition 2.24, it suffices to prove (A7) for only atomic wfs without individual constants. But, hypothesis (a) enables us easily to prove

$$\vdash (y_1 = z_1 \wedge \ldots \wedge y_n = z_n) \Rightarrow (\mathscr{B}(y_1,\ldots,y_n) \Rightarrow \mathscr{B}(z_1,\ldots,z_n))$$

for all variables $y_1,\ldots,y_n$, $z_1,\ldots,z_n$ and any atomic wf $\mathscr{B}(y_1,\ldots,y_n)$ without function letters or individual constants. Hence, it suffices to show:

(*) If $t(x,x)$ is a term without individual constants and $t(x,y)$ comes from $t(x,x)$ by replacing some occurrences of x by y, then $\vdash x = y \Rightarrow t(x,x) = t(x,y)$.†

But (*) can be proved, using hypothesis (b), by induction on the number of function letters in $t(x,x)$, and we leave this as an exercise.

It is easy to see from Proposition 2.25 that, when the language of K has only finitely many predicate and function letters, it is only necessary to verify (A7) for a finite list of special cases (in fact, n wfs for each A_j^n and n wfs for each f_j^n).

Exercises

2.66 Let K_1 be a theory whose language has only $=$ as a predicate letter and no function letters or individual constants. Let its proper axioms be $(\forall x_1)\, x_1 = x_1$, $(\forall x_1)(\forall x_2)(x_1 = x_2 \Rightarrow x_2 = x_1)$ and $(\forall x_1)(\forall x_2)(\forall x_3)(x_1 = x_2 \Rightarrow (x_2 = x_3 \Rightarrow x_1 = x_3))$. Show that K_1 is a theory with equality. [*Hint*: It

†The reader can clarify how (*) is applied by using it to prove the following instance of (A7): $\vdash x = y \Rightarrow (A_1^1(f_1^1(x)) \Rightarrow A_1^1(f_1^1(y)))$. Let $t(x,x)$ be $f_1^1(x)$ and let $t(x,y)$ be $f_1^1(y)$.

suffices to prove that $\vdash x_1 = x_3 \Rightarrow (x_1 = x_2 \Rightarrow x_3 = x_2)$ and $\vdash x_2 = x_3 \Rightarrow (x_1 = x_2 \Rightarrow x_1 = x_3)$.] K_1 is called the *pure first-order theory of equality.*

2.67 Let K_2 be a theory whose language has only $=$ and $<$ as predicate letters and no function letters or individual constants. Let K_2 have the following proper axioms.

(a) $(\forall x_1)\, x_1 = x_1$
(b) $(\forall x_1)(\forall x_2)(x_1 = x_2 \Rightarrow x_2 = x_1)$
(c) $(\forall x_1)(\forall x_2)(\forall x_3)(x_1 = x_2 \Rightarrow (x_2 = x_3 \Rightarrow x_1 = x_3))$
(d) $(\forall x_1)(\exists x_2)(\exists x_3)(x_1 < x_2 \wedge x_3 < x_1)$
(e) $(\forall x_1)(\forall x_2)(\forall x_3)(x_1 < x_2 \wedge x_2 < x_3 \Rightarrow x_1 < x_3)$
(f) $(\forall x_1)(\forall x_2)(x_1 = x_2 \Rightarrow \neg\, x_1 < x_2)$
(g) $(\forall x_1)(\forall x_2)(x_1 < x_2 \vee x_1 = x_2 \vee x_2 < x_1)$
(h) $(\forall x_1)(\forall x_2)(x_1 < x_2 \Rightarrow (\exists x_3)(x_1 < x_3 \wedge x_3 < x_2))$

Using Proposition 2.25, show that K_2 is a theory with equality. K_2 is called the *theory of densely ordered sets with neither first nor last element.*

2.68 Let K be any theory with equality. Prove the following.

(a) $\vdash x_1 = y_1 \wedge \ldots \wedge x_n = y_n \Rightarrow t(x_1, \ldots, x_n) = t(y_1, \ldots, y_n)$, where $t(y_1, \ldots, y_n)$ arises from the term $t(x_1, \ldots, x_n)$ by substitution of $y_1, \ldots, y_n$ for $x_1, \ldots, x_n$, respectively.

(b) $\vdash x_1 = y_1 \wedge \ldots \wedge x_n = y_n \Rightarrow (\mathscr{B}(x_1, \ldots, x_n) \Leftrightarrow \mathscr{B}(y_1, \ldots, y_n))$, where $\mathscr{B}(y_1, \ldots, y_n)$ is obtained by substituting $y_1, \ldots, y_n$ for one or more occurrences of $x_1, \ldots, x_n$, respectively, in the wf $\mathscr{B}(x_1, \ldots, x_n)$, and $y_1, \ldots, y_n$ are free for $x_1, \ldots, x_n$, respectively, in the wf $\mathscr{B}(x_1, \ldots, x_n)$.

Examples.
(In the literature, 'elementary' is sometimes used instead of 'first-order'.)

1. *Elementary theory* G *of groups*: predicate letter $=$, function letter f_1^2, and individual constant a_1. We abbreviate $f_1^2\,(t,s)$ by $t + s$ and a_1 by 0. The proper axioms are the following.
   (a) $x_1 + (x_2 + x_3) = (x_1 + x_2) + x_3$
   (b) $x_1 + 0 = x_1$
   (c) $(\forall x_1)(\exists x_2) x_1 + x_2 = 0$
   (d) $x_1 = x_1$
   (e) $x_1 = x_2 \Rightarrow x_2 = x_1$
   (f) $x_1 = x_2 \Rightarrow (x_2 = x_3 \Rightarrow x_1 = x_3)$
   (g) $x_1 = x_2 \Rightarrow (x_1 + x_3 = x_2 + x_3 \wedge x_3 + x_1 = x_3 + x_2)$
   That G is a theory with equality follows easily from Proposition 2.25. If one adds to the axioms the following wf:
   (h) $x_1 + x_2 = x_2 + x_1$
   the new theory is called the *elementary theory of abelian groups.*
2. *Elementary theory* F *of fields*: predicate letter $=$, function letters f_1^2 and f_2^2, and individual constants a_1 and a_2. Abbreviate $f_1^2(t,s)$ by

$t+s$, $f_2^2(t,s)$ by $t \cdot s$, and a_1 and a_2 by 0 and 1. As proper axioms, take (a)–(h) of Example 1 plus the following.

(i) $x_1 = x_2 \Rightarrow (x_1 \cdot x_3 = x_2 \cdot x_3 \wedge x_3 \cdot x_1 = x_3 \cdot x_2)$
(j) $x_1 \cdot (x_2 \cdot x_3) = (x_1 \cdot x_2) \cdot x_3$
(k) $x_1 \cdot (x_2 + x_3) = (x_1 \cdot x_2) + (x_1 \cdot x_3)$
(l) $x_1 \cdot x_2 = x_2 \cdot x_1$
(m) $x_1 \cdot 1 = x_1$
(n) $x_1 \neq 0 \Rightarrow (\exists x_2)\, x_1 \cdot x_2 = 1$
(o) $0 \neq 1$

F is a theory with equality. Axioms (a)–(m) define the elementary theory R_C of commutative rings with unit. If we add to F the predicate letter A_2^2, abbreviate $A_2^2(t,s)$ by $t < s$, and add axioms (e), (f) and (g) of Exercise 2.67, as well as $x_1 < x_2 \Rightarrow x_1 + x_3 < x_2 + x_3$ and $x_1 < x_2 \wedge 0 < x_3 \Rightarrow x_1 \cdot x_3 < x_2 \cdot x_3$, then the new theory $F_<$ is called the *elementary theory of ordered fields.*

Exercise

2.69 (a) What formulas must be derived in order to use Proposition 2.25 to conclude that the theory G of Example 1 is a theory with equality?
(b) Show that the axioms (d)–(f) of equality mentioned in Example 1 can be replaced by (d) and
(f′): $x_1 = x_2 \Rightarrow (x_3 = x_2 \Rightarrow x_1 = x_3)$.

One often encounters theories K in which $=$ may be defined; that is, there is a wf $\mathscr{E}(x,y)$ with two free variables x and y, such that, if we abbreviate $\mathscr{E}(t,s)$ by $t = s$, then axioms (A6) and (A7) are provable in K. We make the convention that, if t and s are terms that are not free for x and y, respectively, in $\mathscr{E}(x,y)$, then, by suitable changes of bound variables (see Exercise 2.48), we replace $\mathscr{E}(x,y)$ by a logically equivalent wf $\mathscr{E}^*(x,y)$ such that t and s are free for x and y, respectively, in $\mathscr{E}^*(x,y)$; then $t = s$ is to be the abbreviation of $\mathscr{E}^*(t,s)$. Proposition 2.23 and analogues of Propositions 2.24 and 2.25 hold for such theories. There is no harm in extending the term *theory with equality* to cover such theories.

In theories with equality it is possible to define in the following way phrases that use the expression 'There exists one and only one x such that...'.

DEFINITION

$(\exists_1 x)\mathscr{B}(x)$ for $(\exists x)\mathscr{B}(x) \wedge (\forall x)(\forall y)(\mathscr{B}(x) \wedge \mathscr{B}(y) \Rightarrow x = y)$

In this definition, the new variable y is assumed to be the first variable that does not occur in $\mathscr{B}(x)$. A similar convention is to be made in all other definitions where new variables are introduced.

Exercise

2.70 In any theory with equality, prove the following.

(a) $\vdash (\forall x)(\exists_1 y)\ x = y$

(b) $\vdash (\exists_1 x)\mathscr{B}(x) \Leftrightarrow (\exists x)(\forall y)(x = y \Leftrightarrow \mathscr{B}(y))$

(c) $\vdash (\forall x)(\mathscr{B}(x) \Leftrightarrow \mathscr{C}(x)) \Rightarrow [(\exists_1 x)\mathscr{B}(x) \Leftrightarrow (\exists_1 x)\mathscr{C}(x)]$

(d) $\vdash (\exists_1 x)(\mathscr{B} \vee \mathscr{C}) \Rightarrow ((\exists_1 x)\mathscr{B}) \vee (\exists_1 x)\mathscr{C}$

(e) $\vdash (\exists_1 x)\mathscr{B}(x) \Leftrightarrow (\exists x)(\mathscr{B}(x) \wedge (\forall y)(\mathscr{B}(y) \Rightarrow y = x))$

In any model for a theory K with equality, the relation E in the model corresponding to the predicate letter $=$ is an equivalence relation (by Proposition 2.23). If this relation E is the identity relation in the domain of the model, then the model is said to be *normal.*

Any model M for K can be *contracted* to a normal model M^* for K by taking the domain D^* of M^* to be the set of equivalence classes determined by the relation E in the domain D of M. For a predicate letter A_j^n and for any equivalence classes $[b_1], \ldots, [b_n]$ in D^* determined by elements $b_1, \ldots, b_n$ in D, we let $(A_j^n)^{M^*}$ hold for $([b_1], \ldots, [b_n])$ if and only if $(A_j^n)^M$ holds for $(b_1, \ldots, b_n)$. Notice that it makes no difference which representatives $b_1, \ldots, b_n$ we select in the given equivalence classes because, from (A7), $\vdash x_1 = y_1 \wedge \ldots \wedge x_n = y_n \Rightarrow (A_j^n(x_1, \ldots, x_n) \Leftrightarrow A_j^n(y_1, \ldots, y_n))$. Likewise, for any function letter f_j^n and any equivalence classes $[b_1], \ldots, [b_n]$ in D^*, let $(f_j^n)^{M^*}([b_1], \ldots, [b_n]) = [(f_j^n)^M(b_1, \ldots, b_n)]$. Again note that this is independent of the choice of the representatives $b_1, \ldots, b_n$, since, from (A7), we can prove $\vdash x_1 = y_1 \wedge \ldots \wedge x_n = y_n \Rightarrow f_j^n(x_1, \ldots, x_n) = f_j^n(y_1, \ldots, y_n)$. For any individual constant a_i let $(a_i)^{M^*} = [(a_i)^M]$. The relation E^* corresponding to $=$ in the model M^* is the identity relation in D^*: $E^*([b_1], [b_2])$ if and only if $E(b_1, b_2)$, that is, if and only if $[b_1] = [b_2]$. Now one can easily prove by induction the following lemma: If $s = (b_1, b_2, \ldots)$ is a denumerable sequence of elements of D, and $s' = ([b_1], [b_2], \ldots)$ is the corresponding sequence of equivalence classes, then a wf $\mathscr{B}$ is satisfied by s in M if and only if $\mathscr{B}$ is satisfied by s' in M^*. It follows that, for any wf $\mathscr{B}$, $\mathscr{B}$ is true for M if and only if $\mathscr{B}$ is true for M^*. Hence, because M is a model of K, M^* is a normal model of K.

PROPOSITION 2.26 (EXTENSION OF PROPOSITION 2.17)

(Gödel, 1930) Any consistent theory with equality K has a finite or denumerable normal model.

Proof

By Proposition 2.17, K has a denumerable model M. Hence, the contraction of M to a normal model yields a finite or denumerable normal model M^*

because the set of equivalence classes in a denumerable set D is either finite or denumerable.

COROLLARY 2.27 (EXTENSION OF THE SKOLEM–LÖWENHEIM THEOREM)

Any theory with equality K that has an infinite normal model M has a denumerable normal model.

Proof

Add to K the denumerably many new individual constants $b_1, b_2, \ldots$ together with the axioms $b_i \neq b_j$ for $i \neq j$. Then the new theory K′ is consistent. If K′ were inconsistent, there would be a proof in K′ of a contradiction $\mathscr{C} \wedge \neg\mathscr{C}$, where we may assume that $\mathscr{C}$ is a wf of K. But this proof uses only a finite number of the new axioms: $b_{i_1} \neq b_{j_1}, \ldots, b_{i_n} \neq b_{j_n}$. Now, M can be extended to a model $M^{\#}$ of K plus the axioms $b_{i_1} \neq b_{j_1}, \ldots, b_{i_n} \neq b_{j_n}$; in fact, since M is an infinite normal model, we can choose interpretations of $b_{i_1}, b_{j_1}, \ldots, b_{i_n}, b_{j_n}$, so that the wfs $b_{i_1} \neq b_{j_1}, \ldots, b_{i_n} \neq b_{j_n}$ are true. But, since $\mathscr{C} \wedge \neg\mathscr{C}$ is derivable from these wfs and the axioms of K, it would follow that $\mathscr{C} \wedge \neg\mathscr{C}$ is true for $M^{\#}$, which is impossible. Hence, K′ must be consistent. Now, by Proposition 2.26, K′ has a finite or denumerable normal model N. But, since, for $i \neq j$, the wfs $b_i \neq b_j$ are axioms of K′, they are true for N. Thus, the elements in the domain of N that are the interpretations of $b_1, b_2, \ldots$ must be distinct, which implies that the domain of N is infinite and, therefore, denumerable.

Exercises

2.71 We define $(\exists_n x)\mathscr{B}(x)$ by induction on $n \geqslant 1$. The case $n = 1$ has already been taken care of. Let $(\exists_{n+1} x)\mathscr{B}(x)$ stand for
$(\exists y)(\mathscr{B}(y) \wedge (\exists_n x)(x \neq y \wedge \mathscr{B}(x)))$.

(a) Show that $(\exists_n x)\mathscr{B}(x)$ asserts that there are exactly n objects for which $\mathscr{B}$ holds, in the sense that in any normal model for $(\exists_n x)\mathscr{B}(x)$ there are exactly n objects for which the property corresponding to $\mathscr{B}(x)$ holds.

(b) (i) For each positive integer n, write a closed wf $\mathscr{B}_n$ such that $\mathscr{B}_n$ is true in a normal model when and only when that model contains at least n elements.

(ii) Prove that the theory K, whose axioms are those of the pure theory of equality K_1 (see Exercise 2.66), plus the axioms $\mathscr{B}_1, \mathscr{B}_2, \ldots$, is not finitely axiomatizable, that is, there is no theory K′ with a finite number of axioms such that K and K′ have the same theorems.

(iii) For a normal model, state in ordinary English the meaning of $\neg\mathscr{B}_{n+1}$.

(c) Let n be a positive integer and consider the wf $(\mathscr{E}_n)$ $(\exists_n x)x = x$. Let L_n be the theory $\mathrm{K}_1 + \{\mathscr{E}_n\}$, where K_1 is the pure theory of equality.

(i) Show that a normal model M is a model of L_n if and only if there are exactly n elements in the domain of M.

(ii) Define a procedure for determining whether any given sentence is a theorem of L_n and show that L_n is a complete theory.

2.72 (a) Prove that, if a theory with equality K has arbitrarily large finite normal models, then it has a denumerable normal model.

(b) Prove that there is no theory with equality whose normal models are precisely all finite normal interpretations.

2.73 Prove that any predicate calculus with equality is consistent. (A predicate calculus with equality is assumed to have (A1)–(A7) as its only axioms.)

2.74$^{\mathrm{D}}$ Prove the independence of axioms (A1)–(A7) in any predicate calculus with equality.

2.75 If $\mathscr{B}$ is a wf that does not contain the $=$ symbol and $\mathscr{B}$ is provable in a predicate calculus with equality K, show that $\mathscr{B}$ is provable in K without using (A6) or (A7).

2.76$^{\mathrm{D}}$ Show that $=$ can be defined in any theory whose language has only a finite number of predicate letters and no function letters.

2.77 (a)$^{\mathrm{A}}$ Find a non-normal model of the elementary theory of groups G.

(b) Show that any model M of a theory with equality K can be extended to a non-normal model of K. [*Hint*: Use the argument in the proof of the lemma within the proof of Corollary 2.22.]

2.78 Let $\mathscr{B}$ be a wf of a theory with equality. Show that $\mathscr{B}$ is true in every normal model of K if and only if $\vdash_{\mathrm{K}} \mathscr{B}$.

2.79 Write the following as wfs of a theory with equality.

(a) There are at least three moons of Jupiter.

(b) At most two people know everyone in the class.

2.80 If $P(u)$ means *u is a person*, $G(u, v)$ means *u is a grandparent of v*, and $u = v$ means that u and v are identical, translate the following wf into ordinary English:

$$(\forall x)(P(x) \Rightarrow (\exists x_1)(\exists x_2)(\exists x_3)(\exists x_4)(x_1 \neq x_2 \wedge x_1 \neq x_3 \wedge x_1 \neq x_4 \wedge x_2 \neq x_3 \wedge x_2 \neq x_4 \wedge x_3 \neq x_4 \wedge G(x_1, x) \wedge G(x_2, x) \wedge G(x_3, x) \wedge G(x_4, x) \wedge (\forall y)(G(y, x) \Rightarrow y = x_1 \vee y = x_2 \vee y = x_3 \vee y = x_4)))$$

2.81 Consider the wf

$$(*) \quad (\forall x)(\forall y)(\exists z)(z \neq x \wedge z \neq y \wedge A(z)).$$

Show that $(*)$ is true in a normal model M of a theory with equality if and only if there exist in the domain of M at least three things having property $A(z)$.

2.82 Let the language $\mathscr{L}$ have the four predicate letters $=$, P, S and L. Read $u = v$ as *u and v are identical*, $P(u)$ as *u is a point*, $S(u)$ as *u is a line*, and $L(u, v)$ as *u lies on v*. Let the theory of equality $\mathbb{G}$ of *planar incidence geometry* have, in addition to axioms (A1)–(A7), the following non-logical axioms.

(1) $P(x) \Rightarrow \neg S(x)$
(2) $L(x,y) \Rightarrow P(x) \wedge S(y)$
(3) $S(x) \Rightarrow (\exists y)(\exists z)(y \neq z \wedge L(y,x) \wedge L(z,x))$
(4) $P(x) \wedge P(y) \wedge x \neq y \Rightarrow (\exists_1 z)(S(z) \wedge L(x,z) \wedge L(y,z))$
(5) $(\exists x)(\exists y)(\exists z)(P(x) \wedge P(y) \wedge P(z) \wedge \neg \mathscr{C}(x,y,z))$

where $\mathscr{C}(x,y,z)$ is the wf $(\exists u)(S(u) \wedge L(x,u) \wedge L(y,u) \wedge L(z,u))$, which is read as *$x$, y, z are collinear*.

(a) Translate (1)–(5) into ordinary geometric language.
(b) Prove $\vdash_{\mathbb{G}} (\forall u)(\forall v)\ (S(u) \wedge S(v) \wedge u \neq v \Rightarrow\ (\forall x)(\forall y)\ (L(x,u) \wedge L(x,v) \wedge L(y,u) \wedge L(y,v) \Rightarrow x = y))$, and translate this theorem into ordinary geometric language.
(c) Let $R(u,v)$ stand for $S(u) \wedge S(v) \wedge \neg(\exists w)(L(w,u) \wedge L(w,v))$. Read $R(u,v)$ as *u and v are distinct parallel lines*.
  (i) Prove: $\vdash_{\mathbb{G}} R(u,v) \Rightarrow u \neq v$
  (ii) Show that there exists a normal model of $\mathbb{G}$ with a finite domain in which the following sentence is true:

$$(\forall x)(\forall y)(S(x) \wedge P(y) \wedge \neg L(y,x) \Rightarrow (\exists_1 z)(L(y,z) \wedge R(z,x)))$$

(d) Show that there exists a model of $\mathbb{G}$ in which the following sentence is true:

$$(\forall x)(\forall y)(S(x) \wedge S(y) \wedge x \neq y \Rightarrow \neg R(x,y))$$

2.9 DEFINITIONS OF NEW FUNCTION LETTERS AND INDIVIDUAL CONSTANTS

In mathematics, once we have proved, for any $y_1, \ldots, y_n$, the existence of a unique object u that has a property $\mathscr{B}(u, y_1, \ldots, y_n)$, we often introduce a new function letter $f(y_1, \ldots, y_n)$ such that $\mathscr{B}(f(y_1, \ldots, y_n), y_1, \ldots, y_n)$ holds for all $y_1, \ldots, y_n$. In cases where we have proved the existence of a unique object u that satisfies a wf $\mathscr{B}(u)$ and $\mathscr{B}(u)$ contains u as its only free variable, then we introduce a new individual constant b such that $\mathscr{B}(b)$ holds. It is generally acknowledged that such definitions, though convenient, add nothing really new to the theory. This can be made precise in the following manner.

PROPOSITION 2.28

Let K be a theory with equality. Assume that $\vdash_K (\exists_1 u)\mathscr{B}(u, y_1, \ldots, y_n)$. Let $K^\#$ be the theory with equality obtained by adding to K a new function letter f of n arguments and the proper axiom $\mathscr{B}(f(y_1, \ldots, y_n), y_1, \ldots, y_n)$,† as well as all instances of axioms (A1)–(A7) that involve f. Then there is an effective transformation mapping each wf $\mathscr{C}$ of $K^\#$ into a wf $\mathscr{C}^\#$ of K such that:

(a) If f does not occur in $\mathscr{C}$, then $\mathscr{C}^\#$ is $\mathscr{C}$.
(b) $(\neg\mathscr{C})^\#$ is $\neg(\mathscr{C}^\#)$.
(c) $(\mathscr{C} \Rightarrow \mathscr{D})^\#$ is $\mathscr{C}^\# \Rightarrow \mathscr{D}^\#$.
(d) $((\forall x)\mathscr{C})^\#$ is $(\forall x)(\mathscr{C}^\#)$.
(e) $\vdash_{K^\#} (\mathscr{C} \Leftrightarrow \mathscr{C}^\#)$.
(f) If $\vdash_{K^\#} \mathscr{C}$, then $\vdash_K \mathscr{C}^\#$.

Hence, if $\mathscr{C}$ does not contain f and $\vdash_{K^\#} \mathscr{C}$, then $\vdash_K \mathscr{C}$.

Proof

By a *simple f-term* we mean an expression $f(t_1, \ldots, t_n)$ in which $t_1, \ldots, t_n$ are terms that do not contain f. Given an atomic wf $\mathscr{C}$ of $K^\#$, let $\mathscr{C}^*$ be the result of replacing the leftmost occurrence of a simple term $f(t_1, \ldots, t_n)$ in $\mathscr{C}$ by the first variable v not in $\mathscr{C}$ or $\mathscr{B}$. Call the wf $(\exists v)(\mathscr{B}(v, t_1, \ldots, t_n) \wedge \mathscr{C}^*)$ the *f-transform* of $\mathscr{C}$. If $\mathscr{C}$ does not contain f, then let $\mathscr{C}$ be its own f-transform. Clearly, $\vdash_{K^\#} (\exists v)(\mathscr{B}(v, t_1, \ldots, t_n) \wedge \mathscr{C}^*) \Leftrightarrow \mathscr{C}$. (Here, we use $\vdash_K (\exists_1 u)\mathscr{B}(u, y_1, \ldots, y_n)$ and the axiom $\mathscr{B}(f(y_1, \ldots, y_n), y_1, \ldots, y_n)$ of $K^\#$.) Since the f-transform $\mathscr{C}'$ of $\mathscr{C}$ contains one less f than $\mathscr{C}$ and $\vdash_{K^\#} \mathscr{C}' \Leftrightarrow \mathscr{C}$, if we take successive f-transforms, eventually we obtain a wf $\mathscr{C}^\#$ that does not contain f and such that $\vdash_{K^\#} \mathscr{C}^\# \Leftrightarrow \mathscr{C}$. Call $\mathscr{C}^\#$ the *f-less transform* of $\mathscr{C}$. Extend the definition to all wfs of $K^\#$ by letting $(\neg\mathscr{D})^\#$ be $\neg(\mathscr{D}^\#)$, $(\mathscr{D} \Rightarrow \mathscr{E})^\#$ be $\mathscr{D}^\# \Rightarrow \mathscr{E}^\#$, and $((\forall x)\mathscr{D})^\#$ be $(\forall x)\mathscr{D}^\#$. Properties (a)–(e) of Proposition 2.28 are then obvious. To prove property (f), it suffices, by property (e), to show that, if $\mathscr{C}$ does not contain f and $\vdash_{K^\#} \mathscr{C}$, then $\vdash_{K^\#} \mathscr{C}$. We may assume that $\mathscr{C}$ is a closed wf, since a wf and its closure are deducible from each other.

Assume that M is a model of K. Let M_1 be the normal model obtained by contracting M. We know that a wf is true for M if and only if it is true for M_1. Since $\vdash_K (\exists_1 u)\mathscr{B}(u, y_1, \ldots, y_n)$, then, for any $b_1, \ldots, b_n$ in the domain of M_1, there is a unique c in the domain of M_1 such that $\models_{M_1} \mathscr{B}[c, b_1, \ldots, b_n]$. If we define $f_1(b_1, \ldots, b_n)$ to be c, then, taking f_1 to be the interpretation of the function letter f, we obtain from M_1 a model $M^\#$ of $K^\#$. For the logical axioms of $K^\#$ (including the equality axioms of $K^\#$) are true in any normal

†It is better to take this axiom in the form $(\forall u)(u = f(y_1, \ldots, y_n) \Rightarrow \mathscr{B}(u, y_1, \ldots, y_n))$, since $f(y_1, \ldots, y_n)$ might not be free for u in $\mathscr{B}(u, y_1, \ldots, y_n)$.

interpretation, and the axiom $\mathscr{B}(f(y_1,\ldots,y_n),y_1,\ldots,y_n)$ also holds in $M^{\#}$ by virtue of the definition of f_1. Since the other proper axioms of $K^{\#}$ do not contain f and since they are true for M_1, they are also true for $M^{\#}$. But $\vdash_{K^{\#}} \mathscr{C}$. Therefore, $\mathscr{C}$ is true for $M^{\#}$, but since $\mathscr{C}$ does not contain f, $\mathscr{C}$ is true for M_1 and hence also for M. Thus, $\mathscr{C}$ is true for every model of K. Therefore, by Corollary 2.20(a), $\vdash_K \mathscr{C}$. (In the case where $\vdash_K (\exists_1 u)\mathscr{B}(u)$ and $\mathscr{B}(u)$ contains only u as a free variable, we form $K^{\#}$ by adding a new individual constant b and the axiom $\mathscr{B}(b)$. Then the analogue of Proposition 2.28 follows from practically the same proof as the one just given.)

Exercise

2.83 Find the f-less transforms of the following wfs.

(a) $(\forall x)(\exists y)(A_1^3(x,y,f(x,y_1,\ldots,y_n)) \Rightarrow f(y,x,\ldots,x) = x)$
(b) $A_1^1(f(y_1,\ldots,y_{n-1},f(y_1,\ldots,y_n))) \wedge (\exists x)A_1^2(x,f(y_1,\ldots,y_n))$

Note that Proposition 2.28 also applies when we have introduced several new symbols $f_1,\ldots,f_m$ because we can assume that we have added each f_i to the theory already obtained by the addition of $f_1,\ldots,f_{i-1}$; then m successive applications of Proposition 2.28 are necessary. The resulting wf $\mathscr{C}^{\#}$ of K can be considered an $(f_1,\ldots,f_m)$-free transform of $\mathscr{C}$ into the language of K.

Examples

1. In the elementary theory G of groups, one can prove $(\exists_1 y)\, x + y = 0$. Then introduce a new function f of one argument, abbreviate $f(t)$ by $(-t)$, and add the new axiom $x + (-x) = 0$. By Proposition 2.28, we now are not able to prove any wf of G that we could not prove before. Thus, the definition of $(-t)$ adds no really new power to the original theory.
2. In the elementary theory F of fields, one can prove that $(\exists_1 y)((x \neq 0 \wedge x \cdot y = 1) \vee (x = 0 \text{ and } y = 0))$. We then introduce a new function letter g of one argument, abbreviate $g(t)$ by t^{-1}, and add the axiom $(x \neq 0 \wedge x \cdot x^{-1} = 1) \vee (x = 0 \text{ and } x^{-1} = 0)$, from which one can prove $x \neq 0 \Rightarrow x \cdot x^{-1} = 1$.

From Proposition 2.28 we can see that, in theories with equality, only predicate letters are needed; function letters and individual constants are dispensable. If f_j^n is a function letter, we can replace it by a new predicate letter A_k^{n+1} if we add the axiom $(\exists_1 u)A_k^{n+1}(u,y_1,\ldots,y_n)$. An individual constant is to be replaced by a new predicate letter A_k^1 if we add the axiom $(\exists_1 u)A_k^1(u)$.

Example

In the elementary theory G of groups, we can replace + and 0 by predicate letters A_1^3 and A_1^1 if we add the axioms $(\forall x_1)(\forall x_2)\ (\exists_1 x_3)A_1^3(x_1,x_2,x_3)$ and $(\exists_1 x_1)A_1^1(x_1)$, and if we replace axioms (a), (b), (c) and (g) by the following:

(a′) $A_1^3(x_2,x_3,u) \wedge A_1^3(x_1,u,v) \wedge A_1^3(x_1,x_2,w) \wedge A_1^3(w,x_3,y) \Rightarrow v = y$
(b′) $A_1^1(y) \wedge A_1^3(x,y,z) \Rightarrow z = x$
(c′) $(\exists y)(\forall u)(\forall v)(A_1^1(u) \wedge A_1^3(x,y,v) \Rightarrow v = u)$
(g′) $[x_1 = x_2 \wedge A_1^3(x_1,y,z) \wedge A_1^3(x_2,y,u) \wedge A_1^3(y,x_1,v) \wedge A_1^3(y,x_2,w)]$
$\Rightarrow z = u \wedge v = w$

Notice that the proof of Proposition 2.28 is highly non-constructive, since it uses semantical notions (model, truth) and is based upon Corollary 2.20(a), which was proved in a non-constructive way. Constructive syntactical proofs have been given for Proposition 2.28 (see Kleene, 1952, § 74), but, in general, they are quite complex.

Descriptive phrases of the kind 'the u such that $\mathscr{B}(u,y_1,\ldots,y_n)$' are very common in ordinary language and in mathematics. Such phrases are called *definite descriptions*. We let $\iota u(\mathscr{B}(u,y_1,\ldots,y_n))$ denote the unique object u such that $\mathscr{B}(u,y_1,\ldots,y_n)$ if there is such a unique object. If there is no such unique object, either we may let $\iota u(\mathscr{B}(u,y_1,\ldots,y_n))$ stand for some fixed object, or we may consider it meaningless. (For example, we may say that the phrases 'the present king of France' and 'the smallest integer' are meaningless or we may arbitrarily make the convention that they denote 0.) There are various ways of incorporating these ι-terms in formalized theories, but since in most cases the same results are obtained by using new function letters or individual constants as above, and since they all lead to theorems similar to Proposition 2.28, we shall not discuss them any further here. For details, see Hilbert and Bernays (1934) and Rosser (1939; 1953).

2.10 PRENEX NORMAL FORMS

A wf $(Q_1y_1)\ldots(Q_ny_n)\mathscr{B}$, where each (Q_i,y_i) is either $(\forall y_i)$ or $(\exists y_i)$, y_i is different from y_j for $i \neq j$, and $\mathscr{B}$ contains no quantifiers, is said to be in *prenex normal form*. (We include the case $n = 0$, when there are no quantifiers at all.) We shall prove that, for every wf, we can construct an equivalent prenex normal form.

LEMMA 2.29

In any theory, if y is not free in $\mathscr{D}$, and $\mathscr{C}(x)$ and $\mathscr{C}(y)$ are similar, then the following hold.

(a) $\vdash ((\forall x)\mathscr{C}(x) \Rightarrow \mathscr{D}) \Leftrightarrow (\exists y)(\mathscr{C}(y) \Rightarrow \mathscr{D})$
(b) $\vdash ((\exists x)\mathscr{C}(x) \Rightarrow \mathscr{D}) \Leftrightarrow (\forall y)(\mathscr{C}(y) \Rightarrow \mathscr{D})$
(c) $\vdash (\mathscr{D} \Rightarrow (\forall x)\mathscr{C}(x)) \Leftrightarrow (\forall y)(\mathscr{D} \Rightarrow \mathscr{C}(y))$
(d) $\vdash (\mathscr{D} \Rightarrow (\exists x)\mathscr{C}(x)) \Leftrightarrow (\exists y)(\mathscr{D} \Rightarrow \mathscr{C}(y))$
(e) $\vdash \neg(\forall x)\mathscr{C} \Leftrightarrow (\exists x)\neg\mathscr{C}$
(f) $\vdash \neg(\exists x)\mathscr{C} \Leftrightarrow (\forall x)\neg\mathscr{C}$

Proof

For part (a):

1. $(\forall x)\mathscr{C}(x) \Rightarrow \mathscr{D}$	Hyp
2. $\neg(\exists y)(\mathscr{C}(y) \Rightarrow \mathscr{D})$	Hyp
3. $\neg\neg(\forall y)\neg(\mathscr{C}(y) \Rightarrow \mathscr{D})$	2, abbreviation
4. $(\forall y)\neg(\mathscr{C}(y) \Rightarrow \mathscr{D})$	3, negation elimination
5. $(\forall y)(\mathscr{C}(y) \wedge \neg\mathscr{D})$	4, tautology, Proposition 2.9(c)
6. $\mathscr{C}(y) \wedge \neg\mathscr{D}$	5, rule A4
7. $\mathscr{C}(y)$	6, conjunction elimination
8. $(\forall y)\mathscr{C}(y)$	7, Gen
9. $(\forall x)\mathscr{C}(x)$	8, Lemma 2.11, biconditional elimination
10. $\mathscr{D}$	1, 9, MP
11. $\neg\mathscr{D}$	6, conjunction elimination
12. $\mathscr{D} \wedge \neg\mathscr{D}$	10, 11, conjunction introduction
13. $(\forall x)\mathscr{C}(x) \Rightarrow \mathscr{D}$, $\neg(\exists y)(\mathscr{C}(y) \Rightarrow \mathscr{D}) \vdash \mathscr{D} \wedge \neg\mathscr{D}$	1–12
14. $(\forall x)\mathscr{C}(x) \Rightarrow \mathscr{D}$ $\vdash (\exists y)(\mathscr{C}(y) \Rightarrow \mathscr{D})$	1–13, proof by contradiction
15. $\vdash (\forall x)\mathscr{C}(x) \Rightarrow \mathscr{D}$ $\Rightarrow (\exists y)(\mathscr{C}(y) \Rightarrow \mathscr{D})$	1–14, Corollary 2.6

The converse is proven in the following manner.

1. $(\exists y)(\mathscr{C}(y) \Rightarrow \mathscr{D})$	Hyp
2. $(\forall x)\mathscr{C}(x)$	Hyp
3. $\mathscr{C}(b) \Rightarrow \mathscr{D}$	1, rule C
4. $\mathscr{C}(b)$	2, rule A4
5. $\mathscr{D}$	3, 4, MP
6. $(\exists y(\mathscr{C}(y) \Rightarrow \mathscr{D})$, $(\forall x)\mathscr{C}(x) \vdash_C \mathscr{D}$	1–5
7. $(\exists y)(\mathscr{C}(y) \Rightarrow \mathscr{D})$, $(\forall x)\mathscr{C}(x) \vdash \mathscr{D}$	6, Proposition 2.10
8. $\vdash (\exists y)(\mathscr{C}(y) \Rightarrow \mathscr{D})$ $\Rightarrow ((\forall x)\mathscr{C}(x) \Rightarrow \mathscr{D})$	1–7, Corollary 2.6 twice

Part (a) follows from the two proofs above by biconditional introduction. Parts (b)–(f) are proved easily and left as an exercise. (Part (f) is trivial, and (e) appeared as Exercise 2.33(a); (c) and (d) follow easily from (b) and (a), respectively.)

Lemma 2.29 allows us to move interior quantifiers to the front of a wf. This is the essential process in the proof of the following proposition.

PROPOSITION 2.30

There is an effective procedure for transforming any wf $\mathscr{B}$ into a wf $\mathscr{C}$ in prenex normal form such that $\vdash \mathscr{B} \Leftrightarrow \mathscr{C}$.

Proof

We describe the procedure by induction on the number k of occurrences of connectives and quantifiers in $\mathscr{B}$. (By Exercise 2.32(a, b), we may assume that the quantified variables in the prefix that we shall obtain are distinct.) If $k = 0$, then let $\mathscr{C}$ be $\mathscr{B}$ itself. Assume that we can find a corresponding $\mathscr{C}$ for all wfs with $k < n$, and assume that $\mathscr{B}$ has n occurrences of connectives and quantifiers.

Case 1. If $\mathscr{B}$ is $\neg\mathscr{D}$, then, by inductive hypothesis, we can construct a wf $\mathscr{E}$ in prenex normal form such that $\vdash \mathscr{D} \Leftrightarrow \mathscr{E}$. Hence, $\vdash \neg\mathscr{D} \Leftrightarrow \neg\mathscr{E}$ by biconditional negation. Thus, $\vdash \mathscr{B} \Leftrightarrow \neg\mathscr{E}$, and, by applying parts (e) and (f) of Lemma 2.29 and the replacement theorem (Proposition 2.9(b)), we can find a wf $\mathscr{C}$ in prenex normal form such that $\vdash \neg\mathscr{E} \Leftrightarrow \mathscr{C}$. Hence, $\vdash \mathscr{B} \Leftrightarrow \mathscr{C}$.

Case 2. If $\mathscr{B}$ is $\mathscr{D} \Rightarrow \mathscr{E}$, then, by inductive hypothesis, we can find wfs $\mathscr{D}_1$ and $\mathscr{E}_1$ in prenex normal form such that $\vdash \mathscr{D} \Leftrightarrow \mathscr{D}_1$ and $\vdash \mathscr{E} \Leftrightarrow \mathscr{E}_1$. Hence, by a suitable tautology and MP, $\vdash (\mathscr{D} \Rightarrow \mathscr{E}) \Leftrightarrow (\mathscr{D}_1 \Rightarrow \mathscr{E}_1)$, that is, $\vdash \mathscr{B} \Leftrightarrow (\mathscr{D}_1 \Rightarrow \mathscr{E}_1)$. Now, applying parts (a)–(d) of Lemma 2.29 and the replacement theorem, we can move the quantifiers in the prefixes of $\mathscr{D}_1$ and $\mathscr{E}_1$ to the front, obtaining a wf $\mathscr{C}$ in prenex normal form such that $\vdash \mathscr{B} \Leftrightarrow \mathscr{C}$.

Case 3. If $\mathscr{B}$ is $(\forall x)\mathscr{D}$, then, by inductive hypothesis, there is a wf $\mathscr{D}_1$ in prenex normal form such that $\vdash \mathscr{D} \Leftrightarrow \mathscr{D}_1$; hence, $\vdash \mathscr{B} \Leftrightarrow (\forall x)\mathscr{D}_1$ by Gen, Lemma 2.8, and MP. But $(\forall x)\mathscr{D}_1$ is in prenex normal form.

Examples

1. Let $\mathscr{B}$ be $(\forall x)(A_1^1(x) \Rightarrow (\forall y)(A_2^2(x,y) \Rightarrow \neg(\forall z)A_3^2(y,z)))$. By part (e) of Lemma 2.29: $(\forall x)(A_1^1(x) \Rightarrow (\forall y)[A_2^2(x,y) \Rightarrow (\exists z)\neg A_3^2(y,z)])$.
   By part (d): $(\forall x)(A_1^1(x) \Rightarrow (\forall y)(\exists u)[A_2^2(x,y) \Rightarrow \neg A_3^2(y,u)])$.
   By part (c): $(\forall x)(\forall v)(A_1^1(x) \Rightarrow (\exists u)[A_2^2(x,v) \Rightarrow \neg A_3^2(v,u)])$.
   By part (d): $(\forall x)(\forall v)(\exists w)(A_1^1(x) \Rightarrow (A_2^2(x,v) \Rightarrow \neg A_3^2(v,w)))$.
   Changing bound variables: $(\forall x)(\forall y)(\exists z)(A_1^1(x) \Rightarrow (A_2^2(x,y) \Rightarrow \neg A_3^2(y,z)))$.
2. Let $\mathscr{B}$ be $A_1^2(x,y) \Rightarrow (\exists y)[A_1^1(y) \Rightarrow ([(\exists x)A_1^1(x)] \Rightarrow A_2^1(y))]$.
   By part (b): $A_1^2(x,y) \Rightarrow (\exists y)(A_1^1(y) \Rightarrow (\forall u)[A_1^1(u) \Rightarrow A_2^1(y)])$.
   By part (c): $A_1^2(x,y) \Rightarrow (\exists y)(\forall v)(A_1^1(y) \Rightarrow [A_1^1(v) \Rightarrow A_2^1(y)])$.
   By part (d): $(\exists w)(A_1^2(x,y) \Rightarrow (\forall v)[A_1^1(w) \Rightarrow (A_1^1(v) \Rightarrow A_2^1(w))])$.
   By part (c): $(\exists w)(\forall z)(A_1^2(x,y) \Rightarrow [A_1^1(w) \Rightarrow (A_1^1(z) \Rightarrow A_2^1(w))])$.

Exercise

2.84 Find prenex normal forms equivalent to the following wfs.
(a) $[(\forall x)(A_1^1(x) \Rightarrow A_1^2(x,y))] \Rightarrow ([(\exists y)A_1^1(y)] \Rightarrow (\exists z)A_1^2(y,z))$
(b) $(\exists x)A_1^2(x,y) \Rightarrow (A_1^1(x) \Rightarrow \neg(\exists u)A_1^2(x,u))$

A predicate calculus in which there are no function letters or individual constants and in which, for any positive integer n, there are infiitely many predicate letters with n arguments, will be called a *pure predicate calculus*. For pure predicate calculi we can find a very simple prenex normal form theorem. A wf in prenex normal form such that all existential quantifiers (if any) precede all universal quantifiers (if any) is said to be in *Skolem normal form*.

PROPOSITION 2.31

In a pure predicate calculus, there is an effective procedure assigning to each wf $\mathscr{B}$ another wf $\mathscr{S}$ in Skolem normal form such that $\vdash \mathscr{B}$ if and only if $\vdash \mathscr{S}$ (or, equivalently, by Gödel's completeness theorem, such that $\mathscr{B}$ is logically valid if and only if $\mathscr{S}$ is logically valid).

Proof

First we may assume that $\mathscr{B}$ is a closed wf, since a wf is provable if and only if its closure is provable. By Proposition 2.30 we may also assume that $\mathscr{B}$ is in prenex normal form. Let the *rank* r of $\mathscr{B}$ be the number of universal quantifiers in $\mathscr{B}$ that precede existential quantifiers. By induction on the rank, we shall describe the process for finding Skolem normal forms. Clearly, when the rank is 0, we already have the Skolem normal form. Let us assume that we can construct Skolem normal forms when the rank is less than r, and let r be the rank of $\mathscr{B}$. $\mathscr{B}$ can be written as follows: $(\exists y_1)\ldots(\exists y_n)(\forall u)\mathscr{C}(y_1,\ldots,y_n,u)$, where $\mathscr{C}(y_1,\ldots,y_n,u)$ has only $y_1,\ldots,y_n,u$ as its free variables. Let A_j^{n+1} be the first predicate letter of $n+1$ arguments that does not occur in $\mathscr{B}$. Construct the wf

$$(\mathscr{B}_1) \qquad (\exists y_1)\ldots(\exists y_n)([(\forall u)(\mathscr{C}(y_1,\ldots,y_n,u) \Rightarrow A_j^{n+1}(y_1,\ldots,y_n,u))] \Rightarrow (\forall u)A_j^{n+1}(y_1,\ldots,y_n,u))$$

Let us show that $\vdash \mathscr{B}$ if and only if $\vdash \mathscr{B}_1$. Assume $\vdash \mathscr{B}_1$. In the proof of $\mathscr{B}_1$, replace all occurrences of $A_j^{n+1}(z_1,\ldots,z_n,w)$ by $\mathscr{C}^*(z_1,\ldots,z_n,w))$, where $\mathscr{C}^*$ is obtained from $\mathscr{C}$ by replacing all bound variables having free occurrences in the proof by new variables not occurring in the proof. The result is a proof of

$$(\exists y_1)\dots(\exists y_n)(((\forall u)(\mathscr{C}(y_1,\dots,y_n,u) \Rightarrow \mathscr{C}^*(y_1,\dots y_n,u))) \Rightarrow (\forall u)\mathscr{C}^*(y_1,\dots,y_n,u))$$

($\mathscr{C}^*$ was used instead of $\mathscr{C}$ so that applications of axiom (A4) would remain applications of the same axiom.) Now, by changing the bound variables back again, we see that

$$\vdash (\exists y_1)\dots(\exists y_n)[(\forall u)(\mathscr{C}(y_1,\dots,y_n,u) \Rightarrow \mathscr{C}(y_1,\dots,y_n,u)) \Rightarrow (\forall u)\mathscr{C}(y_1,\dots,y_n,u)]$$

Since $\vdash (\forall u)(\mathscr{C}(y_1,\dots,y_n,u) \Rightarrow \mathscr{C}(y_1,\dots,y_n,u))$, we obtain, by the replacement theorem, $\vdash (\exists y_1)\dots(\exists y_n)(\forall u)\mathscr{C}(y_1,\dots,y_n,u)$, that is, $\vdash \mathscr{B}$. Conversely, assume that $\vdash \mathscr{B}$. By rule C, we obtain $(\forall u)\mathscr{C}(b_1,\dots,b_n,u)$. But, $\vdash (\forall u)\mathscr{D} \Rightarrow ((\forall u)(\mathscr{D} \Rightarrow \mathscr{E}) \Rightarrow (\forall u)\mathscr{E})$ (see Exercise 2.27 (a)) for any wfs $\mathscr{D}$ and $\mathscr{E}$. Hence, $\vdash_C (\forall u)(\mathscr{C}(b_1,\dots,b_n,u) \Rightarrow A_j^{n+1}(b_1,\dots,b_n,u)) \Rightarrow (\forall u)A_j^{n+1}(b_1,\dots,b_n,u)$. So, by rule E4, $\vdash_C (\exists y_1)\dots(\exists y_n)$ $([(\forall u)(\mathscr{C}(b_1,\dots b_n,u) \Rightarrow A_j^{n+1}(y_1,\dots,y_n,u))] \Rightarrow (\forall u)A_j^{n+1}(y_1,\dots,y_n,u))$, that is, $\vdash_C \mathscr{B}_1$. By Proposition 2.10, $\vdash \mathscr{B}_1$. A prenex normal form of $\mathscr{B}_1$ has the form $\mathscr{B}_2$: $(\exists y_1)\dots(\exists y_n)(\exists u)(Q_1z_1)\dots(Q_sz_s)(\forall v)\mathscr{G}$, where $\mathscr{G}$ has no quantifiers and $(Q_1,z_1)\dots(Q_sz_s)$ is the prefix of $\mathscr{C}$. [In deriving the prenex normal form, first, by Lemma 2.29(a), we pull out the first $(\forall u)$, which changes to $(\exists u)$; then we pull out of the first conditional the quantifiers in the prefix of $\mathscr{C}$. By Lemma 2.29(a, b), this exchanges existential and universal quantifiers, but then we again pull these out of the second conditional of $\mathscr{B}_1$, which brings the prefix back to its original form. Finally, by Lemma 2.29(c), we bring the second $(\forall u)$ out to the prefix, changing it to a new quantifier $(\forall v)$.] Clearly, $\mathscr{B}_2$ has rank one less than the rank of $\mathscr{B}$ and, by Proposition 2.30, $\vdash \mathscr{B}_1 \Leftrightarrow \mathscr{B}_2$. But, $\vdash \mathscr{B}$ if and only if $\vdash \mathscr{B}_1$. Hence, $\vdash \mathscr{B}$ if and only if $\vdash \mathscr{B}_2$. By inductive hypothesis, we can find a Skolem normal form for $\mathscr{B}_2$, which is also a Skolem normal form for $\mathscr{B}$.

Example

$\mathscr{B}$: $(\forall x)(\forall y)(\exists z)\mathscr{C}(x,y,z)$, where $\mathscr{C}$ contains no quantifiers

$\mathscr{B}_1$: $(\forall x)((\forall y)(\exists z)\mathscr{C}(x,y,z) \Rightarrow A_j^1(x)) \Rightarrow (\forall x)A_j^1(x)$, where A_j^1 is not in $\mathscr{C}$.

We obtain the prenex normal form of $\mathscr{B}_1$:

$$(\exists x)([(\forall y)(\exists z)\mathscr{C}(x,y,z) \Rightarrow A_j^1(x)] \Rightarrow (\forall x)A_j^1(x)) \qquad 2.29(a)$$

$$(\exists x)((\exists y)[(\exists z)\mathscr{C}(x,y,z) \Rightarrow A_j^1(x)] \Rightarrow (\forall x)A_j^1(x)) \qquad 2.29(a)$$

$$(\exists x)((\exists y)(\forall z)[\mathscr{C}(x,y,z) \Rightarrow A_j^1(x)] \Rightarrow (\forall x)A_j^1(x)) \qquad 2.29(b)$$

$$(\exists x)(\forall y)[(\forall z)(\mathscr{C}(x,y,z) \Rightarrow A_j^1(x)) \Rightarrow (\forall x)A_j^1(x)] \qquad 2.29(b)$$

$$(\exists x)(\forall y)(\exists z)([\mathscr{C}(x,y,z) \Rightarrow A_j^1(x)] \Rightarrow (\forall x)A_j^1(x)) \qquad 2.29(a)$$

$$(\exists x)(\forall y)(\exists z)(\forall v)[(\mathscr{C}(x,y,z) \Rightarrow A_j^1(x)) \Rightarrow A_j^1(v)] \qquad 2.29(c)$$

We repeat this process again: Let $\mathscr{D}(x,y,z,v)$ be $(\mathscr{C}(x,y,z) \Rightarrow A_j^1(x)) \Rightarrow A_j^1(v)$. Let A_k^2 not occur in $\mathscr{D}$. Form:

$$(\exists x)([(\forall y)[(\exists z)(\forall v)(\mathscr{D}(x,y,z,v)) \Rightarrow A_k^2(x,y)]] \Rightarrow (\forall y)A_k^2(x,y))$$

$$(\exists x)(\exists y)[[(\exists z)(\forall v)(\mathscr{D}(x,y,z,v)) \Rightarrow A_k^2(x,y)] \Rightarrow (\forall y)A_k^2(x,y)] \qquad 2.29(a)$$

$$(\exists x)(\exists y)(\exists z)(\forall v)([\mathscr{D}(x,y,z,v) \Rightarrow A_k^2(x,y)] \Rightarrow (\forall y)A_k^2(x,y)) \qquad 2.29(a,b)$$

$$(\exists x)(\exists y)(\exists z)(\forall v)(\forall w)([\mathscr{D}(x,y,z,v) \Rightarrow A_k^2(x,y)] \Rightarrow A_k^2(x,w)) \qquad 2.29(c)$$

Thus, a Skolem normal form of $\mathscr{B}$ is:

$$(\exists x)(\exists y)(\exists z)(\forall v)(\forall w)([[(\mathscr{C}(x,y,z) \Rightarrow A_j^1(x)) \Rightarrow A_j^1(v)) \Rightarrow A_k^2(x,y)] \Rightarrow A_k^2(x,w))$$

Exercises

2.85 Find Skolem normal forms for the following wfs.

(a) $\neg(\exists x)A_1^1(x) \Rightarrow (\forall u)(\exists y)(\forall x)A_1^3(u,x,y)$

(b) $(\forall x)(\exists y)(\forall u)(\exists v)A_1^4(x,y,u,v)$

2.86 Show that there is an effective procedure that gives, for each wf $\mathscr{B}$ of a pure predicate calculus, another wf $\mathscr{D}$ of this calculus of the form $(\forall y_1)\ldots(\forall y_n)(\exists z_1)\ldots(\exists z_m)\mathscr{C}$, such that $\mathscr{C}$ is quantifier-free, $n, m \geqslant 0$, and $\mathscr{B}$ is satisfiable if and only if $\mathscr{D}$ is satisfiable. [*Hint:* Apply Proposition 2.31 to $\neg\mathscr{B}$.]

2.87 Find a Skolem normal form $\mathscr{S}$ for $(\forall x)(\exists y)A_1^2(x,y)$ and show that it is not the case that $\vdash \mathscr{S} \Leftrightarrow (\forall x)(\exists y)A_1^2(x,y)$. Hence, a Skolem normal form for a wf $\mathscr{B}$ is not necessarily logically equivalent to $\mathscr{B}$, in contradistinction to the prenex normal form given by Proposition 2.30.

2.11 ISOMORPHISM OF INTERPRETATIONS. CATEGORICITY OF THEORIES

We shall say that an interpretation M of some language $\mathscr{L}$ is *isomorphic* with an interpretation M* of $\mathscr{L}$ if and only if there is a one-one correspondence g (called an isomorphism) of the domain D of M with the domain D^* of M* such that:

1. For any predicate letter A_j^n of $\mathscr{L}$ and for any $b_1,\ldots,b_n$ in D, $\models_{\mathrm{M}} A_j^n[b_1,\ldots,b_n]$ if and only if $\models_{\mathrm{M}^*} A_j^n[g(b_1),\ldots,g(b_n)]$.
2. For any function letter f_j^n of $\mathscr{L}$ and for any $b_1,\ldots,b_n$ in D, $g((f_j^n)^{\mathrm{M}}(b_1,\ldots,b_n)) = (f_j^n)^{\mathrm{M}^*}(g(b_1),\ldots,g(b_n))$.
3. For any individual constant a_j of $\mathscr{L}$, $g((a_j)^{\mathrm{M}}) = (a_j)^{\mathrm{M}^*}$.

The notation $\mathrm{M} \approx \mathrm{M}^*$ will be used to indicate that M is isomorphic with M*. Notice that, if $\mathrm{M} \approx \mathrm{M}^*$, then the domains of M and M* must be of the same cardinality.

PROPOSITION 2.32

If g is an isomorphism of M with M*, then:

(a) for any wf $\mathscr{B}$ of $\mathscr{L}$, any sequence $s = (b_1, b_2, \ldots)$ of elements of the domain D of M, and the corresponding sequence $g(s) = (g(b_1), g(b_2), \ldots)$, s satisfies $\mathscr{B}$ in M if and only if $g(s)$ satisfies $\mathscr{B}$ in M*;
(b) hence, $\models_{\mathrm{M}} \mathscr{B}$ if and only if $\models_{\mathrm{M}^*} \mathscr{B}$.

Proof

Part (b) follows directly from part (a). The proof of part (a) is by induction on the number of connectives and quantifiers in $\mathscr{B}$ and is left as an exercise.

From the definition of isomorphic interpretations and Proposition 2.32 we see that isomorphic interpretations have the same 'structure' and, thus, differ in no essential way.

Exercises

2.88 Prove that, if M is an interpretation with domain D and D^* is a set that has the same cardinality as D, then one can define an interpretation M* with domain D^* such that M is isomorphic with M*.

2.89 Prove the following: (a) M is isomorphic with M. (b) If M_1 is isomorphic with M_2, then M_2 is isomorphic with M_1. (c) If M_1 is isomorphic with M_2 and M_2 is isomorphic with M_3, then M_1 is isomorphic with M_3.

A theory with equality K is said to be $\mathfrak{m}-$*categorical*, where $\mathfrak{m}$ is a cardinal number, if and only if: any two normal models of K of cardinality $\mathfrak{m}$ are isomorphic; and K has at least one normal model of cardinality $\mathfrak{m}$ (see Loś, 1954c).

Examples

1. Let K^2 be the pure theory of equality K_1 (see page 98) to which has been added axiom (E2): $(\exists x_1)(\exists x_2)(x_1 \neq x_2 \wedge (\forall x_3)\ (x_3 = x_1 \vee x_3 = x_2))$. Then K^2 is 2-categorical. Every normal model of K^2 has exactly two elements. More generally, define (En) to be:

$$(\exists x_1)\ldots(\exists x_n)\left(\bigwedge_{1\leqslant i<j\leqslant n} x_i \neq x_j \wedge (\forall y)(y = x_1 \vee \ldots \vee y = x_n)\right)$$

where $\bigwedge_{1\leqslant i<j\leqslant n} x_i \neq x_j$ is the conjunction of all wfs $x_i \neq x_j$ with $1 \leqslant i < j \leqslant n$. Then, if K^n is obtained from K_1 by adding (En) as an axiom, K^n is n-categorical, and every normal model of K^n has exactly n elements.
2. The theory K_2 (see page 98) of densely ordered sets with neither first nor last element is $\aleph_0-$categorical (see Kamke, 1950, p. 71: every denumer-

able normal model of K_2 is isomorphic with the model consisting of the set of rational numbers under their natural ordering). But one can prove that K_2 is not $\mathfrak{m}$–categorical for any $\mathfrak{m}$ different from $\aleph_0$.

Exercises

2.90^A Find a theory with equality that is not $\aleph_0$–categorical but is $\mathfrak{m}$–categorical for all $\mathfrak{m} > \aleph_0$. [*Hint:* Consider the theory G_C of abelian groups (see page 98). For each integer n, let ny stand for the term $(y+y)+\ldots+y$ consisting of the sum of n ys. Add to G_C the axioms $(\mathscr{B}_n)$: $(\forall x)(\exists_1 y)(ny = x)$ for all $n \geqslant 2$. The new theory is the theory of uniquely divisible abelian groups. Its normal models are essentially vector spaces over the field of rational numbers. However, any two vector spaces over the rational numbers of the same non-denumerable cardinality are isomorphic, and there are denumerable vector spaces over the rational numbers that are not isomorphic (see Bourbaki, 1947).]

2.91^A Find a theory with equality that is $\mathfrak{m}$–categorical for all infinite cardinals $\mathfrak{m}$. [*Hint:* Add to the theory G_C of abelian groups the axiom $(\forall x_1)(2x_1 = 0)$. The normal models of this theory are just the vector spaces over the field of integers modulo 2. Any two such vector spaces of the same cardinality are isomorphic (see Bourbaki, 1947).]

2.92 Show that the theorems of the theory K^n in Example 1 above are precisely the set of all wfs of K^n that are true in all normal models of cardinality n.

2.93^A Find two non-isomorphic densely ordered sets of cardinality $2^{\aleph_0}$ with neither first nor last element. (This shows that the theory K_2 of Example 2 is not $2^{\aleph_0}$–categorical.)

Is there a theory with equality that is $\mathfrak{m}$–categorical for some non-countable cardinal $\mathfrak{m}$ but not $\mathfrak{n}$–categorical for some other non-countable cardinal $\mathfrak{n}$? In Example 2 we found a theory that is only $\aleph_0$-categorical; in Exercise 2.90 we found a theory that is $\mathfrak{m}$–categorical for all infinite $\mathfrak{m} > \aleph_0$ but not $\aleph_0$–categorical, and in Exercise 2.91, a theory that is $\mathfrak{m}$–categorical for all infinite $\mathfrak{m}$. The elementary theory G of groups is not $\mathfrak{m}$–categorical for any infinite $\mathfrak{m}$. The problem is whether these four cases exhaust all the possibilities. That this is so was proved by Morley (1965).

2.12 GENERALIZED FIRST-ORDER THEORIES. COMPLETENESS AND DECIDABILITY†

If, in the definition of the notion of first-order language, we allow a non-countable number of predicate letters, function letters, and individual

†Presupposed in parts of this section is a slender acquaintance with ordinal and cardinal numbers (see Chapter 4; or Kamke, 1950; or Sierpinski, 1958).

constants, we arrive at the notion of a *generalized first-order language*. The notions of *interpretation* and *model* extend in an obvious way to a generalized first-order language. A *generalized first-order theory* in such a language is obtained by taking as proper axioms any set of wfs of the language. Ordinary first-order theories are special cases of generalized first-order theories. The reader may easily check that all the results for first-order theories, through Lemma 2.12, hold also for generalized first-order theories without any changes in the proofs. Lemma 2.13 becomes Lemma 2.13′: if the set of symbols of a generalized theory K has cardinality $\aleph_\alpha$, then the set of expressions of K also can be well-ordered and has cardinality $\aleph_\alpha$. (First, fix a well-ordering of the symbols of K. Second, order the expressions by their length, which is some positive integer, and then stipulate that if e_1 and e_2 are two distinct expressions of the same length k, amd j is the first place in which they differ, then e_1 *precedes* e_2 if the jth symbol of e_1 precedes the jth symbol of e_2 according to the given well-ordering of the symbols of K.) Now, under the same assumption as for Lemma 2.13′, Lindenbaum's Lemma 2.14′ can be proved for generalized theories much as before, except that all the enumerations (of the wfs $\mathscr{B}_i$ and of the theories J_i) are transfinite, and the proof that J is consistent and complete uses transfinite induction. The analogue of Henkin's Proposition 2.17 runs as follows.

PROPOSITION 2.33

If the set of symbols of a consistent generalized theory K has cardinality $\aleph_\alpha$, then K has a model of cardinality $\aleph_\alpha$.

Proof

The original proof of Lemma 2.15 is modified in the following way. Add $\aleph_\alpha$ new individual constants $b_1, b_2, \ldots, b_\lambda, \ldots$. As before, the new theory K_0 is consistent. Let $F_1(x_{i_1}), \ldots, F_\lambda(x_{i_\lambda}), \ldots (\lambda < \omega_\alpha)$ be a sequence consisting of all wfs of K_0 with exactly one free variable. Let (S_λ) be the sentence $(\exists x_{i_\lambda})\neg F_\lambda(x_{i_\lambda}) \Rightarrow \neg F_\lambda(b_{j_\lambda})$, where the sequence $b_{j_1}, b_{j_2}, \ldots b_{j_\lambda}, \ldots$ of distinct individual constants is chosen so that b_{j_λ} does not occur in $F_\beta(x_{i_\beta})$ for $\beta \leqslant \lambda$. The new theory K_∞, obtained by adding all the wfs (S_λ) as axioms, is proved to be consistent by a transfinite induction analogous to the inductive proof in Lemma 2.15. K_∞ is a scapegoat theory that is an extension of K and contains $\aleph_\alpha$ closed terms. By the extended Lindenbaum Lemma 2.14′, K_∞ can be extended to a consistent, complete scapegoat theory J with $\aleph_\alpha$ closed terms. The same proof as in Lemma 2.16 provides a model M of J of cardinality $\aleph_\alpha$.

COROLLARY 2.34

(a) If the set of symbols of a consistent generalized theory with equality K has cardinality $\aleph_\alpha$, then K has a normal model of cardinality less than or equal to $\aleph_\alpha$.

(b) If, in addition, K has an infinite normal model (or if K has arbitrarily large finite normal models), then K has a normal model of any cardinality $\aleph_\beta \geqslant \aleph_\alpha$.

(c) In particular, if K is an ordinary theory with equality (i.e., $\aleph_\alpha = \aleph_0$) and K has an infinite normal model (or if K has arbitrarily large finite normal models), then K has a normal model of any cardinality $\aleph_\beta (\beta \geqslant 0)$.

Proof

(a) The model guaranteed by Proposition 2.33 can be contracted to a normal model consisting of equivalence classes in a set of cardinality $\aleph_\alpha$. Such a set of equivalence classes has cardinality less than or equal to $\aleph_\alpha$.

(b) Assume $\aleph_\beta \geqslant \aleph_\alpha$. Let $b_1, b_2, \ldots$ be a set of new individual constants of cardinality $\aleph_\beta$, and add the axioms $b_\lambda \neq b_\mu$ for $\lambda \neq \mu$. As in the proof of Corollary 2.27, this new theory is consistent and so, by (a), has a normal model of cardinality less than or equal to $\aleph_\beta$ (since the new theory has $\aleph_\beta$ new symbols). But, because of the axioms $b_\lambda \neq b_\mu$, the normal model has exactly $\aleph_\beta$ elements.

(c) This is a special case of (b).

Exercise

2.94 If the set of symbols of a predicate calculus with equality K has cardinality $\aleph_\alpha$, prove that there is an extension K′ of K (with the same symbols as K) such that K′ has normal model of cardinality $\aleph_\alpha$, but K′ has no normal model of cardinality less than $\aleph_\alpha$.

From Lemma 2.12 and Corollary 2.34(a, b), it follows easily that, if a generalized theory with equality K has $\aleph_\alpha$ symbols, is $\aleph_\beta$-categorical for some $\beta \geqslant \alpha$, and has no finite models, then K is complete, in the sense that, for any closed wf $\mathscr{B}$, either $\vdash_K \mathscr{B}$ or $\vdash_K \neg\mathscr{B}$ (Vaught, 1954). If not $\vdash_K \mathscr{B}$ and not-$\vdash_K \neg\mathscr{B}$, then the theories $K' = K + \{\neg\mathscr{B}\}$ and $K'' = K + \{\mathscr{B}\}$ are consistent by Lemma 2.12, and so, by Corollary 2.34(a), there are normal models M′ and M″ of K′ and K″, respectively, of cardinality less than or equal to $\aleph_\alpha$. Since K has no finite models, M′ and M″ are infinite. Hence, by Corollary 2.34(b), there are normal models N′ and N″ of K′ and K″, respectively, of cardinality $\aleph_\beta$. By the $\aleph_\beta$-categoricity of K, N′ and N″ must be

isomorphic. But, since $\neg\mathscr{B}$ is true in N′ and $\mathscr{B}$ is true in N″, this is impossible by Proposition 2.32(b). Therefore, either $\vdash_K \mathscr{B}$ or $\vdash_K \neg\mathscr{B}$.

In particular, if K is an ordinary theory with equality that has no finite models and is $\aleph_\beta$-categorical for some $\beta \geqslant 0$, then K is complete. As an example, consider the theory K_2 of densely ordered sets with neither first nor last element (see page 98). K_2 has no finite models and is $\aleph_0$-cateogrical.

If an ordinary theory K is axiomatic (i.e., one can effectively decide whether any wf is an axiom) and complete, then K is decidable, that is, there is an effective procedure to determine whether any given wf is a theorem. To see this, remember (see page 86) that if a theory is axiomatic, one can effectively enumerate the theorems. Any wf $\mathscr{B}$ is provable if and only if its closure is provable. Hence, we may confine our attention to closed wfs $\mathscr{B}$. Since K is complete, either $\mathscr{B}$ is a theorem or $\neg\mathscr{B}$ is a theorem, and, therefore, one or the other will eventually turn up in our enumeration of theorems. This provides an effective test for theoremhood. Notice, that if K is inconsistent, then every wf is a theorem and there is an obvious decision procedure; if K is consistent, then not both $\mathscr{B}$ and $\neg\mathscr{B}$ can show up as theorems and we need only wait until one or the other appears.

If an ordinary axiomatic theory with equality K has no finite models and is $\aleph_\beta$-categorical for some $\beta \geqslant 0$, then, by what we have proved, K is decidable. In particular, the theory K_2 discussed above is decidable.

In certain cases, there is a more direct method of proving completeness or decidability. Let us take as an example the theory K_2 of densely ordered sets with neither first nor last element. Langford (1927) has given the following procedure for K_2. Consider any closed wf $\mathscr{B}$. By proposition 2.30, we can assume that $\mathscr{B}$ is in prenex normal form $(Q_1 y_1)\ldots(Q_n y_n)\mathscr{C}$, where $\mathscr{C}$ contains no quantifiers. If $(Q_n y_n)$ is $(\forall y_n)$, replace $(\forall y_n)\mathscr{C}$ by $\neg(\exists y_n)\neg\mathscr{C}$. In all cases, then, we have, at the right side of the wf, $(\exists y_n)\mathscr{D}$, where $\mathscr{D}$ has no quantifiers. Any negation $x \neq y$ can be replaced by $x < y \vee y < x$, and $\neg(x < y)$ can be replaced by $x = y \vee y < x$. Hence, all negation signs can be eliminated from $\mathscr{D}$. We can now put $\mathscr{D}$ into disjunctive normal form, that is, a disjunction of conjunctions of atomic wfs (see Exercise 1.42). Now $(\exists y_n)(\mathscr{D}_1 \vee \mathscr{D}_2 \vee \ldots \vee \mathscr{D}_k)$ is equivalent to $(\exists y_n)\mathscr{D}_1 \vee (\exists y_n)\mathscr{D}_2 \vee \ldots \vee (\exists y_n)\mathscr{D}_k$. Consider each $(\exists y_n)\mathscr{D}_i$ separately. $\mathscr{D}_i$ is a conjunction of atomic wfs of the form $t < s$ and $t = s$. If $\mathscr{D}_i$ does not contain y_n, just erase $(\exists y_n)$. Note that, if a wf $\mathscr{E}$ does not contain y_n, then $(\exists y_n)(\mathscr{E} \wedge \mathscr{F})$ may be replaced by $\mathscr{E} \wedge (\exists y_n)\mathscr{F}$. Hence, we are reduced to the consideration of $(\exists y_n)\mathscr{F}$, where $\mathscr{F}$ is a conjunction of atomic wfs of the form $t < s$ or $t = s$, each of which contains y_n. Now, if one of the conjuncts is $y_n = z$ for some z different from y_n, then replace in $\mathscr{F}$ all occurrences of y_n by z and erase $(\exists y_n)$. If we have $y_n = y_n$ alone, then just erase $(\exists y_n)$. If we have $y_n = y_n$ as one conjunct among others, then erase $y_n = y_n$. If $\mathscr{F}$ has a conjunct $y_n < y_n$, then replace all of $(\exists y_n)\mathscr{F}$ by $y_n < y_n$. If $\mathscr{F}$ consists of $y_n < z_n \wedge \ldots \wedge y_n < z_j \wedge u_1 < y_n \wedge \ldots \wedge u_m < y_n$, then replace $(\exists y_n)\mathscr{F}$ by the conjunction of all the

wfs $u_i < z_p$ for $1 \leqslant i \leqslant m$ and $1 \leqslant p \leqslant j$. If all the us or all the z_ps are missing, replace $(\exists y_n)\mathscr{F}$ by $y_n = y_n$. This exhausts all possibilities and, in every case, we have replaced $(\exists y_n)\mathscr{F}$ by a wf containing no quantifiers, that is, we have eliminated the quantifier $(\exists y_n)$. We are left with $(Q_1 y_1)\ldots(Q_{n-1} y_{n-1})\mathscr{G}$, where $\mathscr{G}$ contains no quantifiers. Now we apply the same procedure successively to $(Q_{n-1} y_{n-1}), \ldots, (Q_1 y_1)$. Finally we are left with a wf without quantifiers, built up of wfs of the form $x = x$ and $x < x$. If we replace $x = x$ by $x = x \Rightarrow x = x$ and $x < x$ by $\neg(x = x \Rightarrow x = x)$, the result is either an instance of a tautology or the negation of such an instance. Hence, by Proposition 2.1, either the result or its negation is provable. Now, one can easily check that all the replacements we have made in this whole reduction procedure applied to $\mathscr{B}$ have been replacements of wfs $\mathscr{H}$ by other wfs $\mathscr{U}$ such that $\vdash_K \mathscr{H} \Leftrightarrow \mathscr{U}$. Hence, by the replacement theorem, if our final result $\mathscr{R}$ is provable, then so is the original wf $\mathscr{B}$, and, if $\neg\mathscr{R}$ is provable, then so is $\neg\mathscr{B}$. Thus, K_2 is complete and decidable.

The method used in this proof, the successive elimination of existential quantifiers, has been applied to other theories. It yields a decision procedure (see Hilbert and Bernays, 1934, § 5) for the pure theory of equality K_1 (see page 98). It has been applied by Tarski (1951) to prove the completeness and decidability of elementary algebra (i.e., of the theory of real-closed fields; see van der Waerden, 1949) and by Szmielew (1955) to prove the decidability of the theory G_C of abelian groups.

Exercises

2.95 (Henkin, 1955) If an ordinary theory with equality K is finitely axiomatizable and $\aleph_\alpha$-categorical for some α, prove that K is decidable.

2.96 (a) Prove the decidability of the pure theory K_1 of equality.
(b) Give an example of a theory with equality that is $\aleph_\alpha$-categorical for some α, but is incomplete.

Mathematical applications

1. Let F be the elementary theory of fields (see page 98). We let n stand for the term $1 + 1 + \ldots + 1$, consisting of the sum of n 1s. Then the assertion that a field has characteristic p can be expressed by the wf $\mathscr{C}_p$: $p = 0$. A field has characteristic 0 if and only if it does not have characteristic p for any prime p. Then for any closed wf $\mathscr{B}$ of F that is true for all fields of characteristic 0, there is a prime number q such that $\mathscr{B}$ is true for all fields of characteristic greater than or equal to q. To see this, notice that, if F_0 is obtained from F by adding as axioms $\neg\mathscr{C}_2, \neg\mathscr{C}_3, \ldots, \neg\mathscr{C}_p, \ldots$ (for all primes p), the normal models of F_0 are the fields of characteristic 0. Hence, by Exercise 2.77, $\vdash_{F_0} \mathscr{B}$. But then, for some finite set of new axioms

$\neg\mathscr{C}_{q_1}, \neg\mathscr{C}_{q_2}, \ldots, \neg\mathscr{C}_{q_n}$, we have $\neg\mathscr{C}_{q_1}, \neg\mathscr{C}_{q_2}, \ldots, \neg\mathscr{C}_{q_n} \vdash_F \mathscr{B}$. Let q be a prime greater than all $q_1, \ldots, q_n$, In every field of characteristic greater than or equal to q, the wfs $\neg\mathscr{C}_{q_1}, \neg\mathscr{C}_{q_2}, \ldots, \neg\mathscr{C}_{q_n}$ are true; hence, $\mathscr{B}$ is also true. (Other applications in algebra may be found in A. Robinson (1951) and Cherlin (1976).)

2. A *graph* may be considered as a set with a symmetric binary relation R (i.e., the relation that holds between two vertices if and only if they are connected by an edge). Call a graph k-colourable if and only if the graph can be divided into k disjoint (possibly empty) sets such that no two elements in the same set are in the relation R. (Intuitively, these sets correspond to k colours, each colour being painted on the points in the corresponding set, with the proviso that two points connected by an edge are painted different colours.) Notice that any subgraph of a k-colourable graph is k-colourable. Now we can show that, if every finite subgraph of a graph $\mathscr{G}$ is k-colourable, and if $\mathscr{G}$ can be well-ordered, then the whole graph $\mathscr{G}$ is k-colourable. To prove this, construct the following generalized theory with equality K (Beth, 1953). There are two binary predicate letters, $A_1^2(=)$ and A_2^2 (corresponding to the relation R on $\mathscr{G}$); there are k monadic predicate letters $A_1^1, \ldots, A_k^1$ (corresponding to the k subsets into which we hope to divide the graph); and there are individual constants a_c, one for each element c of the graph $\mathscr{G}$. As proper axioms, in addition to the usual assumptions (A6) and (A7), we have the following wfs:

(I) $\neg A_2^2(x,x)$ (irreflexivity of R)

(II) $A_2^2(x,y) \Rightarrow A_2^2(y,x)$ (symmetry of R)

(III) $(\forall x)(A_1^1(x) \vee A_2^1(x) \vee \ldots \vee A_k^1(x))$ (division into k classes)

(IV) $(\forall x)\neg(A_i^1(x) \wedge A_j^1(x))$, for $1 \leqslant i < j \leqslant k$ (disjointness of the k classes)

(V) $(\forall x)(\forall y)(A_i^1(x) \wedge A_i^1(y) \Rightarrow \neg A_2^2(x,y))$ for $1 \leqslant i \leqslant k$ (two elements of the same class are not in the relation R)

(VI) $a_b \neq a_c$, for any two distinct elements b and c of $\mathscr{G}$

(VII) $A_2^2(a_b, a_c)$, if $R(b,c)$ holds in $\mathscr{G}$

Now, any finite set of these axioms involves only a finite number of the individual constants $a_{c_1}, \ldots, a_{c_n}$, and since the corresponding subgraph $\{c_1, \ldots, c_n\}$ is, by assumption, k-colourable, the given finite set of axioms has a model and is, therefore, consistent. Since any finite set of axioms is consistent, K is consistent. By Corollary 2.34(a), K has a normal model of cardinality less than or equal to the cardinality of $\mathscr{G}$. This model is a k-colourable graph and, by (VI)–(VII), has $\mathscr{G}$ as a subgraph. Hence $\mathscr{G}$ is also k-colourable. (Compare this proof with a standard mathematical proof of the same result by Bruijn and Erdös (1951). Generally, use of the method above replaces complicated applications of Tychonoff's theorem or König's Unendlichkeits lemma.)

Exercises

2.97[A] (Loś, 1954b) A group B is said to be *orderable* if there exists a binary relation R on B that totally orders B such that, if xRy, then $(x+z)R(y+z)$ and $(z+x)R(z+y)$. Show, by a method similar to that used in Example 2 above, that a group B is orderable if and only if every finitely generated subgroup is orderable (if we assume that the set B can be well-ordered).

2.98[A] Set up a theory for algebraically closed fields of characteristic $p\,(\geqslant 0)$ by adding to the theory F of fields the new axioms P_n, where P_n states that every non-constant polynomial of degree n has a root, as well as axioms that determine the characteristic. Show that every wf of F that holds for one algebraically closed field of characteristic 0 holds for all of them. [*Hint*: This theory is $\aleph_\beta$-categorical for $\beta > 0$, is axiomatizable, and has no finite models. See A. Robinson (1952).]

2.99 By ordinary mathematical reasoning, solve the *finite marriage problem.* Given a finite set M of m men and a set N of women such that each man knows only a finite number of women and, for $1 \leqslant k \leqslant m$, any subset of M having k elements knows at least k women of N (i.e., there are at least k women in N who know at least one of the k given men), then it is possible to marry (monogamously) all the men of M to women in N so that every man is married to a women whom he knows. [*Hint* (Halmos and Vaughn, 1950): $m = 1$ is trival. For $m > 1$, use induction, considering the cases: (I) for all k with $1 \leqslant k < m$, every set of k men knows at least $k+1$ women; and (II) for some k with $1 \leqslant k < m$, there is a set of k men knowing exactly k women.] Extend this result to the infinite case, that is, when M is infinite and well-orderable and the assumptions above hold for all finite k. [*Hint*: Construct an appropriate generalized theory with equality, analogous to that in Example 2 above, and use Corollary 2.34(a).]

2.100 Prove that there is no generalized theory with equality K, having one predicate letter $<$ in addition to $=$, such that the normal models of K are exactly those normal interpretations in which the interpretation of $<$ is a well-ordering of the domain of the interpretation.

Let $\mathscr{B}$ be a wf in prenex normal form. If $\mathscr{B}$ is not closed, form its closure instead. Suppose, for example, $\mathscr{B}$ is $(\exists y_1)(\forall y_2)(\forall y_3)(\exists y_4)(\exists y_5)(\forall y_6)$ $\mathscr{C}(y_1, y_2, y_3, y_4, y_5, y_6)$, where $\mathscr{C}$ contains no quantifiers. Erase $(\exists y_1)$ and replace y_1 in $\mathscr{C}$ by a new individual constant b_1: $(\forall y_2)(\forall y_3)(\exists y_4)(\exists y_5)(\forall y_6)$ $\mathscr{C}(b_1, y_2, y_3, y_4, y_5, y_6)$. Erase $(\forall y_2)$ and $(\forall y_3)$, obtaining $(\exists y_4)(\exists y_5)(\forall y_6)$ $\mathscr{C}(b_1, y_2, y_3, y_4, y_5, y_6)$. Now erase $(\exists y_4)$ and replace y_4 in $\mathscr{C}$ by $g(y_2, y_3)$, where g is a new function letter: $(\exists y_5)(\forall y_6)\mathscr{C}(b_1, y_2, y_3, g(y_2, y_3), y_5, y_6)$. Erase $(\exists y_5)$ and replace y_5 by $h(y_2, y_3)$, where h is another new function letter: $(\forall y_6)$ $\mathscr{C}(b_1, y_2, y_3, g(y_2, y_3), h(y_2, y_3), y_6)$. Finally, erase $(\forall y_6)$. The resulting wf $\mathscr{C}(b_1, y_2, y_3, g(y_2, y_3), h(y_2, y_3), y_6)$ contains no quantifiers and will be denoted by $\mathscr{B}^*$. Thus, by introducing new function letters and individual constants, we can eliminate the quantifiers from a wf.

Examples

1. If $\mathscr{B}$ is $(\forall y_1)(\exists y_2)(\forall y_3)(\forall y_4)(\exists y_5)\mathscr{C}(y_1, y_2, y_3, y_4, y_5)$, where $\mathscr{C}$ is quantifier-free, then $\mathscr{B}^*$ is of the form $\mathscr{C}(y_1, g(y_1), y_3, y_4, h(y_1, y_3, y_4))$.
2. If $\mathscr{B}$ is $(\exists y_1)(\exists y_2)(\forall y_3)(\forall y_4)(\exists y_5)\mathscr{C}(y_1, y_2, y_3, y_4, y_5)$, where $\mathscr{C}$ is quantifier-free, then $\mathscr{B}^*$ is of the form $\mathscr{C}(b, c, y_3, y_4, g(y_3, y_4))$.

Notice that $\mathscr{B}^* \vdash \mathscr{B}$, since we can put the quantifiers back by applications of Gen and rule E4. (To be more precise, in the process of obtaining $\mathscr{B}^*$, we drop all quantifiers and, for each existentially quantified variable y_i, we substitute a term $g(z_1, \ldots, z_k)$, where g is a new function letter and $z_1, \ldots, z_k$ are the variables that were universally quantified in the prefix preceding $(\exists y_i)$. If there are no such variables $z_1, \ldots, z_k$, we replace y_i by a new individual constant.)

PROPOSITION 2.35 (SECOND ε-THEOREM)

(Rasiowa, 1956; Hilbert and Bernays, 1939) Let K be a generalized theory. Replace each axiom $\mathscr{B}$ of K by $\mathscr{B}^*$. (The new function letters and individual constants introduced foı one axiom are to be different from those introduced for another axiom.) Let K^* be the generalized theory with the proper axioms $\mathscr{B}^*$. Then:

(a) If $\mathscr{D}$ is a wf of K and $\vdash_{K^*} \mathscr{D}$, then $\vdash_K \mathscr{D}$.
(b) K is consistent if and only if K^* is consistent.

Proof

(a) Let $\mathscr{D}$ be a wf of K such that $\vdash_{K^*} \mathscr{D}$. Consider the ordinary theory K° whose axioms $\mathscr{B}_1, \ldots, \mathscr{B}_n$ are such that $\mathscr{B}_1^*, \ldots, \mathscr{B}_n^*$ are the axioms used in the proof of $\mathscr{D}$. Let $K^{\circ *}$ be the theory whose axioms are $\mathscr{B}_1^*, \ldots, \mathscr{B}_n^*$. Hence $\vdash_{K^{\circ *}} \mathscr{D}$. Assume that M is a denumerable model of K°. We may assume that the domain of M is the set P of positive integers (see Exercise 2.88). Let $\mathscr{B}$ be any axiom of K°. For example, suppose that $\mathscr{B}$ has the form $(\exists y_1)(\forall y_2)(\forall y_3)(\exists y_4)\mathscr{C}(y_1, y_2, y_3, y_4)$, where $\mathscr{C}$ is quantifier-free. $\mathscr{B}^*$ has the form $\mathscr{C}(b, y_2, y_3, g(y_2, y_3))$. Extend the model M step by step in the following way (noting that the domain always remains P); since $\mathscr{B}$ is true for M, $(\exists y_1)(\forall y_2)(\forall y_3)(\exists y_4)\mathscr{C}(y_1, y_2, y_3, y_4)$ is true for M. Let the interpretation b^* of b be the least positive integer y_1 such that $(\forall y_2)(\forall y_3)(\exists y_4)\mathscr{C}(y_1, y_2, y_3, y_4)$ is true for M. Hence, $(\exists y_4)\mathscr{C}(b, y_2, y_3, y_4)$ is true in this extended model. For any positive integers y_2 and y_3, let the interpretation of $g(y_2, y_3)$ be the least positive integer y_4 such that $\mathscr{C}(b, y_2, y_3, y_4)$ is true in the extended model. Hence, $\mathscr{C}(b, y_2, y_3, g(y_2, y_3))$ is true in the extended model. If we do this for all the axioms $\mathscr{B}$ of K°, we obtain a model M^* of $K^{\circ *}$. Since $\vdash_{K^{\circ *}} \mathscr{D}$, $\mathscr{D}$ is true

for M^*. Since M^* differs from M only in having interpretations of the new individual constants and function letters, and since $\mathscr{D}$ does not contain any of those symbols, $\mathscr{D}$ is true for M. Thus, $\mathscr{D}$ is true in every denumerable model of K°. Hence, $\vdash_{K^\circ} \mathscr{D}$, by Corollary 2.20(a). Since the axioms of K° are axioms of K, we have $\vdash_K \mathscr{D}$. (For a constructive proof of an equivalent result, see Hilbert and Bernays (1939).)

b) Clearly, K^* is an extension of K, since $\mathscr{B}^* \vdash \mathscr{B}$. Hence, if K^* is consistent, so is K. Conversely, assume K is consistent. Let $\mathscr{D}$ be any wf of K. If K^* is inconsistent, $\vdash_{K^*} \mathscr{D} \wedge \neg\mathscr{D}$. By (a), $\vdash_K \mathscr{D} \wedge \neg\mathscr{D}$, contradicting the consistency of K.

Let us use the term *generalized completeness theorem* for the proposition that every consistent generalized theory has a model. If we assume that every set can be well-ordered (or, equivalently, the axiom of choice), then the generalized completeness theorem is a consequence of Proposition 2.33.

By the *maximal ideal theorem* (MI) we mean the proposition that every proper ideal of a Boolean algebra can be extended to a maximal ideal.† This is equivalent to the Boolean representation theorem, which states that every Boolean algebra is isomorphic to a Boolean algebra of sets. (Compare Stone (1936). For the theory of Boolean algebras, see Sikorski (1960) or Mendelson (1970).) The usual proofs of the MI theorem use the axiom of choice, but it is a remarkable fact that the MI theorem is equivalent to the generalized completeness theorem, and this equivalence can be proved without using the axiom of choice.

PROPOSITION 2.36

(Loś, 1954a; Rasiowa and Sikorski, 1951; 1952) The generalized completeness theorem is equivalent to the maximal ideal theorem.

Proof

(a) Assume the generalized completeness theorem. Let B be a Boolean algebra. Construct a generalized theory with equality K having the binary function letters $\cup$ and $\cap$, the singulary function letter f_1^1 [we denote $f_1^1(t)$ by $\bar{t}$], predicate letters $=$ and A_1^1, and, for each element b in B, an individual constant a_b. By the complete description of B, we mean the following sentences: (i) $a_b \neq a_c$ if b and c are distinct elements of B; (ii) $a_b \cup a_c = a_d$ if b, c, d are elements of B such that $b \cup c = d$ in B; (iii) $a_b \cap a_c = a_e$ if b, c, e are elements of b such that $b \cap c = e$ in B; and (iv) $\bar{a}_b = a_c$ if b and c are elements of B such that $\bar{b} = c$ in B, where $\bar{b}$ denotes the complement of b. As

†Since $\{0\}$ is a proper ideal of a Boolean algebra, this implies (and is implied by) the proposition that every Boolean algebra has a maximal ideal.

axioms of K we take a set of axioms for a Boolean algebra, axioms (A6) and (A7) for equality, the complete description of B, and axioms asserting that A_1^1 determines a maximal ideal (i.e., $A_1^1(x \cap \bar{x})$, $A_1^1(x) \wedge A_1^1(y) \Rightarrow A_1^1(x \cup y)$, $A_1^1(x) \Rightarrow A_1^1(x \cap y)$, $A_1^1(x) \vee A_1^1(\bar{x})$, and $\neg A_1^1(x \cup \bar{x})$). Now K is consistent, for, if there were a proof in K of a contradiction, this proof would contain only a finite number of the symbols $a_b, a_c, \ldots$—say, $a_{b_1}, \ldots, a_{b_n}$. The elements $b_1, \ldots, b_n$ generate a finite subalgebra B′ of B. Every finite Boolean algebra clearly has a maximal ideal. Hence, B′ is a model for the wfs that occur in the proof of the contradiction, and therefore the contradiction is true in B′, which is impossible. Thus, K is consistent and, by the generalized completeness theorem, K has a model. That model can be contracted to a normal model of K, which is a Boolean algebra A with a maximal ideal I. Since the complete description of B is included in the axioms of K, B is a subalgebra of A, and then $I \bigcap B$ is a maximal ideal in B.

(b) Assume the maximal ideal theorem. Let K be a consistent generalized theory. For each axiom $\mathscr{B}$ of K, form the wf $\mathscr{B}^*$ obtained by constructing a prenex normal form for $\mathscr{B}$ and then eliminating the quantifiers through the addition of new individual constants and function letters (see the example preceding the proof of Proposition 2.35). Let $K^{\#}$ be a new theory having the wfs $\mathscr{B}^*$, plus all instances of tautologies, as its axioms, such that its wfs contain no quantifiers and its rules of inference are modus ponens and a rule of substitution for variables (namely, substitution of terms for variables). Now, $K^{\#}$ is consistent, since the theorems of $K^{\#}$ are also theorems of the consistent K^* of Proposition 2.35. Let B be the Lindenbaum algebra determined by $K^{\#}$ (i.e., for any wfs $\mathscr{C}$ and $\mathscr{D}$, let $\mathscr{C}$ Eq $\mathscr{D}$ mean that $\vdash_{K^{\#}} \mathscr{C} \Leftrightarrow \mathscr{D}$; Eq is an equivalence relation; let $[\mathscr{C}]$ be the equivalence class of $\mathscr{C}$; define $[\mathscr{C}] \cup [\mathscr{D}] = [\mathscr{C} \vee \mathscr{D}]$, $[\mathscr{C}] \cap [\mathscr{D}] = [\mathscr{C} \wedge \mathscr{D}]$, $\overline{[\mathscr{C}]} = [\neg \mathscr{C}]$; under these operations, the set of equivalence classes is a Boolean algebra, called the Lindenbaum algebra of $K^{\#}$). By the maximal ideal theorem, let I be a maximal ideal in B. Define a model M of $K^{\#}$ having the set of terms of $K^{\#}$ as its domain; the individual constants and function letters are their own interpretations, and, for any predicate letter A_j^n, we say that $A_j^n(t_1, \ldots, t_n)$ is true in M if and only if $[A_j^n(t_1, \ldots, t_n)]$ is not in I. One can show easily that a wf $\mathscr{C}$ of $K^{\#}$ is true in M if and only if $[\mathscr{C}]$ is not in I. But, for any theorem $\mathscr{D}$ of $K^{\#}$, $[\mathscr{D}] = 1$, which is not in I. Hence, M is a model for $K^{\#}$. For any axiom $\mathscr{B}$ of K, every substitution instance of $\mathscr{B}^*(y_1, \ldots, y_n)$ is a theorem in $K^{\#}$; therefore, $\mathscr{B}^*(y_1, \ldots, y_n)$ is true for all $y_1, \ldots, y_n$ in the model. It follows easily, by reversing the process through which $\mathscr{B}^*$ arose from $\mathscr{B}$, that $\mathscr{B}$ is true in the model. Hence, M is a model for K.

The maximal ideal theorem (and, therefore, also the generalized completeness theorem) turns out to be strictly weaker than the axiom of choice (see Halpern, 1964).

Exercise

2.101 Show that the generalized completeness theorem implies that every set can be totally ordered (and, therefore, that the axiom of choice holds for any set of non-empty disjoint finite sets).

The natural algebraic structures corresponding to the propositional calculus are Boolean algebras (see Exercise 1.60, and Rosenbloom, 1950, chaps 1 and 2). For first-order theories, the presence of quantifiers introduces more algebraic structure. For example, if K is a first-order theory, then, in the corresponding Lindenbaum algebra B, $[(\exists x)\mathscr{B}(x)] = \Sigma_t[\mathscr{B}(t)]$, where Σ_t indicates the least upper bound in B, and t ranges over all terms of K that are free for x in $\mathscr{B}(x)$. Two types of algebraic structure have been proposed to serve as algebraic counterparts of quantification theory. The first, cylindrical algebras, have been studied extensively by Tarski, Thompson, Henkin, Monk and others (see Henkin, Monk and Tarski, 1971). The other approach is the theory of polyadic algebras, invented and developed by Halmos (1962).

2.13 ELEMENTARY EQUIVALENCE. ELEMENTARY EXTENSIONS

Two interpretations M_1 and M_2 of a generalized first-order language $\mathscr{L}$ are said to be *elementarily equivalent* (written $M_1 \equiv M_2$) if the sentences of $\mathscr{L}$ true for M_1 are the same as the sentences true for M_2. Intuitively, $M_1 \equiv M_2$ if and only if M_1 and M_2 cannot be distinguished by means of the language $\mathscr{L}$. Of course, since $\mathscr{L}$ is a generalized first-order language, $\mathscr{L}$ may have non-denumerably many symbols.

Clearly, (1) $M \equiv M$; (2) if $M_1 \equiv M_2$, then $M_2 \equiv M_1$; (3) if $M_1 \equiv M_2$ and $M_2 \equiv M_3$, then $M_1 \equiv M_3$.

Two models of a complete theory K must be elementarily equivalent, since the sentences true in these models are precisely the sentences provable in K. This applies, for example, to any two densely ordered sets without first or last elements (see page 116).

We already know, by Proposition 2.32(b), that isomorphic models are elementarily equivalent. The converse, however, is not true. Consider, for example, any complete theory K that has an infinite normal model. By Corollary 2.34(b), K has normal models of any infinite cardinality $\aleph_\alpha$. If we take two normal models of K of different cardinality, they are elementarily equivalent but not isomorphic. A concrete example is the complete theory K_2 of densely ordered sets that have neither first nor last element. The rational numbers and the real numbers, under their natural orderings, are elementarily equivalent non-isomorphic models of K_2.

Exercises

2.102 Let K_∞, the theory of infinite sets, consist of the pure theory K_1 of equality plus the axioms $\mathscr{B}_n$, where $\mathscr{B}_n$ asserts that there are at least n elements. Show that any two models of K_∞ are elementarily equivalent (see Exercises 2.66 and 2.96(a)).

2.103^D If M_1 and M_2 are elementarily equivalent normal models and M_1 is finite, prove that M_1 and M_2 are isomorphic.

2.104 Let K be a theory with equality having $\aleph_\alpha$ symbols.

(a) Prove that there are at most $2^{\aleph_\alpha}$ models of K, no two of which are elementarily equivalent.

(b) Prove that there are at most $2^{\aleph_\gamma}$ mutually non-isomorphic models of K of cardinality $\aleph_\beta$, where γ is the maximum of α and β.

2.105 Let M be any infinite normal model of a theory with equality K having $\aleph_\alpha$ symbols. Prove that, for any cardinal $\aleph_\gamma \geqslant \aleph_\alpha$, there is a normal model M^* of K of cardinality $\aleph_\alpha$ such that $M \equiv M^*$.

A model M_2 of a language $\mathscr{L}$ is said to be an *extension* of a model M_1 of $\mathscr{L}$ (written $M_1 \subseteq M_2$)[†] if the following conditions hold:

1. The domain D_1 of M_1 is a subset of the domain D_2 of M_2.
2. For any individual constant c of $\mathscr{L}$, $c^{M_2} = c^{M_1}$, where c^{M_2} and c^{M_1} are the interpretations of c in M_2 and M_1.
3. For any function letter f_j^n of $\mathscr{L}$ and any $b_1, \ldots, b_n$ in D, $(f_j^n)^{M_2}(b_1, \ldots, b_n) = (f_j^n)^{M_1}(b_1, \ldots, b_n)$.
4. For any predicate letter A_j^n of $\mathscr{L}$ and any $b_1, \ldots, b_n$ in D, $\models M_1 A_j^n[b_1, \ldots, b_n]$ if and only if $\models_{M_2} A_j^n[b_1, \ldots, b_n]$.

When $M_1 \subseteq M_2$, one also says that M_1 is a *substructure* (or *submodel*) of M_2.

Examples

1. If $\mathscr{L}$ contains only the predicate letters $=$ and $<$, then the set of rational numbers under its natural ordering is an extension of the set of integers under its natural ordering.
2. If $\mathscr{L}$ is the language of field theory (with the predicate letter $=$, function letters $+$ and $\times$, and individual constants 0 and 1), then the field of real numbers is an extension of the field of rational numbers, the field of rational numbers is an extension of the ring of integers, and the ring of integers is an extension of the 'semiring' of non-negative integers. For any fields F_1 and F_2, $F_1 \subseteq F_2$ if and only if F_1 is a subfield of F_2 in the usual algebraic sense.

[†]The reader will have no occasion to confuse this use of $\subseteq$ with that for the inclusion relation.

Exercises

2.106 Prove:
(a) $M \subseteq M$;
(b) if $M_1 \subseteq M_2$ and $M_2 \subseteq M_3$, then $M_1 \subseteq M_3$;
(c) If $M_1 \subseteq M_2$ and $M_2 \subseteq M_1$, then $M_1 = M_2$.

2.107 Assume $M_1 \subseteq M_2$.
(a) Let $\mathscr{B}(x_1, \ldots, x_n)$ be a wf of the form $(\forall y_1) \ldots (\forall y_m)\mathscr{C}(x_1, \ldots x_n, y_1 \ldots, y_m)$, where $\mathscr{C}$ is quantifier-free. Show that, for any $b_1, \ldots, b_n$ in the domain of M_1, if $\models_{M_2} \mathscr{B}[b_1, \ldots, b_n]$, then $\models_{M_1} \mathscr{B}[b_1, \ldots, b_n]$. In particular, any sentence $(\forall y_1) \ldots (\forall y_m)\mathscr{C}(y_1 \ldots, y_m)$, where $\mathscr{C}$ is quantifier-free, is true in M_1 if it is true in M_2.
(b) Let $\mathscr{B}(x_1, \ldots, x_n)$ be a wf of the form $(\exists y_1) \ldots (\exists y_m)\mathscr{C}(x_1, \ldots x_n, y_1, \ldots, y_m)$, where $\mathscr{C}$ is quantifier-free. Show that, for any $b_1, \ldots, b_n$ in the domain of M_1, if $\models_{M_1} \mathscr{B}[b_1, \ldots, b_n]$,then $\models_{M_2} \mathscr{B}[b_1, \ldots, b_n]$. In particular, any sentence $(\exists y_1) \ldots (\exists y_m)\ \mathscr{C}(y_1 \ldots, y_m)$, where $\mathscr{C}$ is quantifier-free, is true in M_2 if it is true in M_1.

2.108 (a) Let K be the predicate calculus of the language of field theory. Find a model M of K and a non-empty subset X of the domain D of M such that there is no substructure of M having domain X.
(b) If K is a predicate calculus with no individual constants or function letters, show that, if M is a model of K and X is a subset of the domain D of M, then there is one and only one substructure of M having domain X.
(c) Let K be any predicate calculus. Let M be any model of K and let X be any subset of the domain D of M. Let Y be the intersection of the domains of all submodels M^* of M such that X is a subset of the domain D_{M^*} of M^*. Show that there is one and only one submodel of M having domain Y. (This submodel is called the *submodel generated* by X.)

A somewhat stronger relation between interpretations than 'extension' is useful in model theory. Let M_1 and M_2 be models of some language $\mathscr{L}$. We say that M_2 is an *elementary extension* of M_1 (written $M_1 \leqslant_e M_2$) if (1) $M_1 \subseteq M_2$ and (2) for any wf $\mathscr{B}(y_1 \ldots, y_n)$ of $\mathscr{L}$ and for any $b_1, \ldots, b_n$ in the domain D_1 of M_1, $\models_{M_1} \mathscr{B}[b_1, \ldots, b_n]$ if and only if $\models_{M_2} \mathscr{B}[b_1, \ldots, b_n]$. (In particular, for any sentence $\mathscr{B}$ of $\mathscr{L}$, $\mathscr{B}$ is true for M_1 if and only if $\mathscr{B}$ is true for M_2.) When $M_1 \leqslant_e M_2$, we shall also say that M_1 is an *elementary substructure* (or *elementary submodel*) of M_2.

It is obvious that, if $M_1 \leqslant_e M_2$, then $M_1 \subseteq M_2$ and $M_1 \equiv M_2$. The converse is not true, as the following example shows. Let G be the elementary theory of groups (see page 98). G has the predicate letter $=$, function letter $+$, and individual constant 0. Let I be the group of integers and E the group of even integers. Then $E \subseteq I$ and $I \cong E$. (The function g

such that $g(x) = 2x$ for all x in I is an isomorphism of I with E.) Consider the wf $\mathscr{B}(y)$: $(\exists x)(x + x = y)$. Then $\models_I \mathscr{B}[2]$, but not-$\models_E \mathscr{B}[2]$. Thus, I is not an elementary extension of E. (This example shows the stronger result that even assuming $M_1 \subseteq M_2$ and $M_1 \cong M_2$ does not imply $M_1 \leqslant_e M_2$.)

The following theorem provides an easy method for showing that $M_1 \leqslant_e M_2$.

PROPOSITION 2.37 (Tarski and Vaught, 1957)

Let $M_1 \subseteq M_2$. Assume the following condition:

($) For every wf $\mathscr{B}(x_1, \ldots x_k)$ of the form $(\exists y)\mathscr{C}(x_1, \ldots x_k, y)$ and for all $b_1, \ldots, b_k$ in the domain D_1 of M_1, if $\models_{M_2} \mathscr{B}[b_1, \ldots, b_k]$, then there is some d in D_1 such that $\models_{M_2} \mathscr{C}[b_1, \ldots, b_k, d]$.

Then $M_1 \leqslant_e M_2$.

Proof

Let us prove:

(*) $\models_{M_1} \mathscr{D}[b_1, \ldots, b_k]$ if and only if $\models_{M_2} \mathscr{D}[b_1, \ldots, b_k]$ for any wf $\mathscr{D}(x_1, \ldots, x_k)$ and any $b_1, \ldots, b_k$ in D_1.

The proof is by induction on the number m of connectives and quantifiers in $\mathscr{D}$. If $m = 0$, then (*) follows from clause 4 of the definition of $M_1 \subseteq M_2$. Now assume that (*) holds true for all wfs having fewer than m connectives and quantifiers.

Case 1. $\mathscr{D}$ is $\neg\mathscr{E}$. By inductive hypothesis, $\models_{M_1} \mathscr{E}[b_1, \ldots, b_k]$ if and only if $\models_{M_2} \mathscr{E}[b_1, \ldots, b_k]$. Using the fact that not - $\models_{M_1} \mathscr{E}[b_1, \ldots, b_k]$ if and only if $\models_{M_1} \neg\mathscr{E}[b_1, \ldots, b_k]$, and similarly for M_2, we obtain (*).

Case 2. $\mathscr{D}$ is $\mathscr{E} \Rightarrow \mathscr{F}$. By inductive hypothesis, $\models_{M_1} \mathscr{E}[b_1, \ldots, b_k]$ if and only if $\models_{M_2} \mathscr{E}[b_1, \ldots, b_k]$ and similarly for $\mathscr{F}$. (*) then follows easily.

Case 3. $\mathscr{D}$ is $(\exists y)\mathscr{E}(x_1, \ldots x_n, y)$. By inductive hypothesis,

(**) $\models_{M_1} \mathscr{E}[b_1, \ldots, b_k, d]$ if and only if $\models_{M_2} \mathscr{E}[b_1, \ldots, b_k, d]$, for any $b_1, \ldots, b_k, d$ in D_1.

Case 3a. Assume $\models_{M_1} (\exists y)\mathscr{E}(x_1, \ldots x_k, y)[b_1, \ldots, b_k]$ for some $b_1, \ldots, b_k$ in D_1. Then $\models_{M_1} \mathscr{E}[b_1, \ldots, b_k, d]$ for some d in D_1. So, by (**), $\models_{M_2} \mathscr{E}[b_1, \ldots, b_k, d]$. Hence, $\models_{M_2} (\exists y)\mathscr{E}(x_1, \ldots x_k, y)[b_1, \ldots, b_k]$.

Case 3b. Assume $\models_{M_2} (\exists y)\mathscr{E}(x_1, \ldots x_k, y)[b_1, \ldots, b_k]$ for some $b_1, \ldots, b_k$ in D_1. By assumption ($), there exists d in D_1 such that $\models_{M_2} \mathscr{E}[b_1, \ldots, b_k, d]$. Hence, by (**), $\models_{M_1} \mathscr{E}[b_1, \ldots, b_k, d]$ and therefore $\models_{M_1} (\exists y)\mathscr{E}(x_1, \ldots x_k, y)[b_1, \ldots, b_k]$.

This completes the induction proof, since any wf is logically equivalent to a wf that can be built up from atomic wfs by forming negations, conditionals and existential quantifications.

Exercises

2.109 Prove:

(a) $M \leqslant_e M$;
(b) if $M_1 \leqslant_e M_2$ and $M_2 \leqslant_e M_3$, then $M_1 \leqslant_e M_3$;
(c) if $M_1 \leqslant_e M$ and $M_2 \leqslant_e M$ and $M_1 \subseteq M_2$, then $M_1 \leqslant_e M_2$.

2.110 Let K be the theory of totally ordered sets with equality (axioms (a)–(c) and (e)–(g) of Exercise 2.67). Let M_1 and M_2 be the models for K with domains the set of positive integers and the set of non-negative integers, respectively (under their natural orderings in both cases). Prove that $M_1 \subseteq M_2$ and $M_1 \simeq M_2$, but $M_1 \not\leqslant_e M_2$.

Let M be an interpretation of a language $\mathscr{L}$. Extend $\mathscr{L}$ to a language $\mathscr{L}^*$ by adding a new individual constant a_d for every member d of the domain of M. We can extend M to an interpretation of $\mathscr{L}^*$ by taking d as the interpretation of a_d. By the *diagram* of M we mean the set of all true sentences of M of the forms $A_j^n(a_{d_1}, \ldots, a_{d_n})$, $\neg A_j^n(a_{d_1}, \ldots, a_{d_n})$, and $f_j^n(a_{d_1}, \ldots, a_{d_n}) = a_{d_m}$. In particular, $a_{d_1} \neq a_{d_2}$ belongs to the diagram if $d_1 \neq d_2$. By the *complete diagram* of M we mean the set of all sentences of $\mathscr{L}^*$ that are true for M.

Clearly, any model $M^{\#}$ of the complete diagram of M determines an elementary extension $M^{\#\#}$ of M,† and vice versa.

Exercise

2.111 (a) Let M_1 be a denumerable normal model of an ordinary theory K with equality such that every element of the domain of M_1 is the interpretation of some closed term of K.
  (i) Show that, if $M_1 \subseteq M_2$ and $M_1 \equiv M_2$, then $M_1 \leqslant_e M_2$.
  (ii) Prove that there is a denumerable normal elementary extension M_3 of M_1 such that M_1 and M_3 are not isomorphic.

(b) Let K be a predicate calculus with equality having two function letters $+$ and $\times$ and two individual constants 0 and 1. Let M be the standard model of arithmetic with domain the set of natural numbers, and $+, \times, 0$ and 1 having their ordinary meaning. Prove that M has a denumerable normal elementary extension that is not isomorphic to M, that is, there is a denumerable nonstandard model of arithmetic.

†The elementary extension $M^{\#\#}$ of M is obtained from $M^{\#}$ by forgetting about the interpretations of the a_ds.

PROPOSITION 2.38 (UPWARD SKOLEM–LÖWENHEIM–TARSKI THEOREM)

Let K be a theory with equality having $\aleph_\alpha$ symbols, and let M be a normal model of K with domain of cardinality $\aleph_\beta$. Let γ be the maximum of α and β. Then, for any $\delta \geqslant \gamma$, there is a model M^* of cardinality $\aleph_\delta$ such that $M \neq M^*$ and $M \leqslant_e M^*$.

Proof

Add to the complete diagram of M a set of cardinality $\aleph_\delta$ of new individual constants b_τ, together with axioms $b_\tau \neq b_\rho$ for distinct τ and ρ and axioms $b_\tau \neq a_d$ for all individual constants a_d corresponding to members d of the domain of M. This new theory $K^\#$ is consistent, since M can be used as a model for any finite number of axioms of $K^\#$. (If $b_{\tau_1}, \ldots, b_{\tau_k}, a_{d_1}, \ldots, a_{d_m}$ are the new individual constants in these axioms, interpret $b_{\tau_1}, \ldots, b_{\tau_k}$ as distinct elements of the domain of M different from $d_1, \ldots, d_m$.) Hence, by Corollary 2.34 (a), $K^\#$ has a normal model $M^\#$ of cardinality $\aleph_\delta$ such that $M \subseteq M^\#$, $M \neq M^\#$, and $M \leqslant_e M^\#$.

PROPOSITION 2.39 (DOWNWARD SKOLEM–LÖWENHEIM–TARSKI THEOREM)

Let K be a theory having $\aleph_\alpha$ symbols, and let M be a model of K with domain of cardinality $\aleph_\gamma \geqslant \aleph_\alpha$. Assume A is a subset of the domain D of M having cardinality $\mathfrak{n}$, and assume $\aleph_\beta$ is such that $\aleph_\gamma \geqslant \aleph_\beta \geqslant \max(\aleph_\alpha, \mathfrak{n})$. Then there is an elementary submodel M^* of M of cardinality $\aleph_\beta$ and with domain D^* including A.

Proof

Since $\mathfrak{n} \leqslant \aleph_\beta \leqslant \aleph_\gamma$, we can add $\aleph_\beta$ elements of D to A to obtain a larger set B of cardinality $\aleph_\beta$. Consider any subset C of D having cardinality $\aleph_\beta$. For every wf $\mathscr{B}(y_1, \ldots, y_n, z)$ of K, and any $c_1, \ldots, c_n$ in C such that $\models_M (\exists z)\mathscr{B}(y_1, \ldots, y_n, z)[c_1, \ldots, c_n]$, add to C the first element d of D (with respect to some fixed well-ordering of D) such that $\models_M (\exists z)\mathscr{B}[c_1, \ldots, c_n, d]$. Denote the so-enlarged set by $C^\#$. Since K has $\aleph_\alpha$ symbols, there are $\aleph_\alpha$ wfs. Since $\aleph_\alpha \leqslant \aleph_\beta$, there are at most $\aleph_\beta$ new elements in $C^\#$ and, therefore, the cardinality of $C^\#$ is $\aleph_\beta$. Form by induction a sequence of sets $C_0, C_1, \ldots$ by setting $C_0 = B$ and $C_{n+1} = C_n^\#$. Let $D^* = \bigcup_{n \in \omega} C_n$. Then the cardinality of D^* is $\aleph_\beta$. In addition, D^* is closed under all the functions $(f_j^n)^M$. (Assume $d_1, \ldots, d_n$ in D^*. We may assume $d_1, \ldots, d_n$ in C_k for some k. Now $\models_M (\exists z)(f_j^n(x_1, \ldots, x_n) = z)[d_1, \ldots, d_n]$. Hence, $(f_j^n)^M(d_1, \ldots, d_n)$, being the

first and only member d of D such that $\models_{\mathrm{M}} (f_j^n(x_1,\ldots,x_n) = z)[d_1,\ldots,d_n,d]$, must belong to $C_k^{\#} = C_{k+1} \subseteq D^*$.) Similarly, all interpretations $(a_j)^{\mathrm{M}}$ of individual constants are in D^*. Hence, D^* determines a substructure M^* of M. To show that $\mathrm{M}^* \leqslant_{\mathrm{e}} \mathrm{M}$, consider any wf $\mathscr{B}(y_1,\ldots,y_n,z)$ and any $d_1,\ldots,d_n$ in D^* such that $\models_{\mathrm{M}} (\exists z)\mathscr{B}(y_1,\ldots,y_n,z)[d_1,\ldots,d_n]$. There exists C_k such that $d_1,\ldots,d_n$ are in C_k. Let d be the first element of D such that $\models_{\mathrm{M}} \mathscr{B}[d_1,\ldots,d_n,d]$. Then $d \in C_k^{\#} = C_{k+1} \subseteq D^*$. So, by the Tarski–Vaught theorem (Proposition 2.37) $\mathrm{M}^* \leqslant_{\mathrm{e}} \mathrm{M}$.

2.14 ULTRAPOWERS. NON-STANDARD ANALYSIS

By a *filter*[†] on a non-empty set A we mean a set $\mathscr{F}$ of subsets of A such that:

1. $\mathrm{A} \in \mathscr{F}$
2. $\mathrm{B} \in \mathscr{F} \wedge \mathrm{C} \in \mathscr{F} \Rightarrow \mathrm{B} \cap \mathrm{C} \in \mathscr{F}$
3. $\mathrm{B} \in \mathscr{F} \wedge \mathrm{B} \subseteq \mathrm{C} \Rightarrow \mathrm{C} \in \mathscr{F}$

Examples
Let $B \subseteq A$. The set $\mathscr{F}_B = \{C \mid B \subseteq C \subseteq A\}$ is a filter on A. $\mathscr{F}_B$ consists of all subsets of A that include B. Any filter of the form $\mathscr{F}_B$ is called a *principal* filter. In particular, $\mathscr{F}_A = \{A\}$ and $\mathscr{F}_\emptyset = \mathscr{P}(A)$ are principal filters. The filter $\mathscr{P}(A)$ is said to be *improper* and every other filter is said to be *proper*.

Exercises

2.112 Show that a filter $\mathscr{F}$ on A is proper if and only if $\emptyset \notin \mathscr{F}$.
2.113 Show that a filter $\mathscr{F}$ on A is a principal filter if and only if the intersection of all sets in $\mathscr{F}$ is a member of $\mathscr{F}$.
2.114 Prove that every finite filter is a principal filter. In particular, any filter on a finite set A is a principal filter.
2.115 Let A be infinite and let $\mathscr{F}$ be the set of all subsets of A that are complements of finite sets: $\mathscr{F} = \{C|(\exists W)(C = A - W \wedge \mathrm{Fin}(W)\}$, where $\mathrm{Fin}(W)$ means that W is finite. Show that $\mathscr{F}$ is a non-principal filter on A.
2.116 Assume A has cardinality $\aleph_\beta$. Let $\aleph_\alpha \leqslant \aleph_\beta$. Let $\mathscr{F}$ be the set of all subsets of A whose complements have cardinality $< \aleph_\alpha$. Show that $\mathscr{F}$ is a non-principal filter on A.
2.117 A collection $\mathscr{G}$ of sets is said to have the *finite intersection property* if $B_1 \cap B_2 \cap \ldots \cap B_k \neq \emptyset$ for any sets $B_1, B_2, \ldots, B_k$ in $\mathscr{G}$. If $\mathscr{G}$ is a collection of

[†]The notion of a filter is related to that of an ideal. A subset $\mathscr{F}$ of $\mathscr{P}(A)$ is a filter on A if and only if the set $\mathscr{G} = \{A - B \mid B \in \mathscr{F}\}$ of complements of sets in $\mathscr{F}$ is an ideal in the Boolean algebra $\mathscr{P}(A)$. Remember that $\mathscr{P}(A)$ denotes the set of all subsets of A.

subsets of A having the finite intersection property and $\mathscr{H}$ is the set of all finite intersections $B_1 \cap B_2 \cap \ldots \cap B_k$ of sets in $\mathscr{G}$, show that $\mathscr{F} = \{D|(\exists C)(B \in \mathscr{H} \wedge C \subseteq D \subseteq A)\}$ is a proper filter on A.

DEFINITION

A filter $\mathscr{F}$ on a set A is called an *ultrafilter* on A if $\mathscr{F}$ is a maximal proper filter on A, that is, $\mathscr{F}$ is a proper filter on A and there is no proper filter $\mathscr{G}$ on A such that $\mathscr{F} \subset \mathscr{G}$.

Example
Let $d \in A$. The principal filter $\mathscr{F}_d = \{B|d \in B \wedge B \subseteq A\}$ is an ultrafilter on A. Assume that $\mathscr{G}$ is a filter on A such that $\mathscr{F}_d \subset \mathscr{G}$. Let $C \in \mathscr{G} - \mathscr{F}_d$. Then $C \subseteq A$ and $d \notin C$. Hence, $d \in A - C$. Thus, $A - C \in \mathscr{F}_d \subset \mathscr{G}$. Since $\mathscr{G}$ is a filter and C and $A - C$ are both in $\mathscr{G}$, then $\emptyset = C \cap (A - C) \in \mathscr{G}$. Hence, $\mathscr{G}$ is not a proper filter.

Exercises

2.118 Let $\mathscr{F}$ be a proper filter on A and assume that $B \subseteq A$ and $A - B \notin \mathscr{F}$. Prove that there is a proper filter $\mathscr{F}' \supseteq \mathscr{F}$ such that $B \in \mathscr{F}'$.
2.119 Let $\mathscr{F}$ be a proper filter on A. Prove that $\mathscr{F}$ is an ultrafilter on A if and only if, for every $B \subseteq A$, either $B \in \mathscr{F}$ or $A - B \in \mathscr{F}$.
2.120 Let $\mathscr{F}$ be a proper filter on A. Show that $\mathscr{F}$ is an ultrafilter on A if and only if, for all B and C in $\mathscr{P}(A)$, if $B \notin \mathscr{F}$ and $C \notin \mathscr{F}$, then $B \cup C \notin \mathscr{F}$.
2.121 (a) Show that every principal ultrafilter on A is of the form $\mathscr{F}_d = \{B|d \in B \wedge B \subseteq A\}$ for some d in A.
(b) Show that a non-principal ultrafilter on A contains no finite sets.
2.122 Let $\mathscr{F}$ be a filter on A and let $\mathscr{I}$ be the corresponding ideal: $B \in \mathscr{I}$ if and only if $A - B \in \mathscr{F}$. Prove that $\mathscr{F}$ is an ultrafilter on A if and only if $\mathscr{I}$ is a maximal ideal.
2.123 Let X be a chain of proper filters on A, that is, for any B and C in X, eithr $B \subseteq C$ or $C \subseteq B$. Prove that the union $\bigcup X = \{a|(\exists B)(B \in X \wedge a \in B)\}$ is a proper filter on A, and $B \subseteq \bigcup X$ for all B in X.

PROPOSITION 2.40 (ULTRAFILTER THEOREM)

Every proper filter $\mathscr{F}$ on a set A can be extended to an ultrafilter on A[†]

[†]We assume the generalized completeness theorem.

Proof

Let $\mathscr{F}$ be a proper filter on A. Let $\mathscr{I}$ be the corresponding proper ideal: $B \in \mathscr{I}$ if and only if $A - B \in \mathscr{F}$. By Proposition 2.36, every ideal can be extended to a maximal ideal. In particular, $\mathscr{I}$ can be extended to a maximal ideal $\mathscr{H}$. If we let $\mathscr{U} = \{B | A - B \in \mathscr{H}\}$, then $\mathscr{U}$ is easily seen to be an ultrafilter and $\mathscr{F} \subseteq \mathscr{U}$.

Alternatively, the existence of an ultrafilter including $\mathscr{F}$ can be proved easily on the basis of Zorn's lemma. (In fact, consider the set X of all proper filters $\mathscr{F}'$ such that $\mathscr{F} \subseteq \mathscr{F}'$. X is partially ordered by $\subset$, and any $\subset$-chain in X has an upper bound in X, namely, by Exercise 2.123, the union of all filters in the chain. Hence, by Zorn's lemma, there is a maximal element $\mathscr{F}^*$ in X, which is the required ultrafilter.) However, Zorn's lemma is equivalent to the axiom of choice, which is a stronger assumption than the generalized completeness theorem.

COROLLARY 2.41

If A is an infinite set, there exists a non-principal ultrafilter on A.

Proof

Let $\mathscr{F}$ be the filter on A consisting of all complements $A - B$ of finite subsets B of A (see Exercise 2.115). By Proposition 2.40, there is an ultrafilter $\mathscr{U} \supseteq \mathscr{F}$. Assume $\mathscr{U}$ is a principal ultrafilter. By Exercise 2.121(a), $\mathscr{U} = \mathscr{F}_d$ for some $d \in A$. Then $A - \{d\} \in \mathscr{F} \subseteq \mathscr{U}$. Also, $\{d\} \in \mathscr{U}$. Hence, $\emptyset = \{d\} \cap (A - \{d\}) \in \mathscr{U}$, contradicting the fact that an ultrafilter is proper.

Reduced direct products

We shall now study an important way of constructing models. Let K be any predicate calculus with equality. Let J be a non-empty set and, for each j in J, let M_j be some normal model of K. In other words, consider a function F assigning to each j in J some normal model. We denote $F(j)$ by M_j.

Let $\mathscr{F}$ be a filter on J. For each j in J, let D_j denote the domain of the model M_j. By the Cartesian product $\Pi_{j \in J} D_j$ we mean the set of all functions f with domain J such that $f(j) \in D_j$ for all j in J. If $f \in \Pi_{j \in J} D_j$, we shall refer to $f(j)$ as the jth *component* of f. Let us define a binary relation $=_{\mathscr{F}}$ in $\Pi_{j \in J} D_j$ as follows:

$$f =_{\mathscr{F}} g \text{ if and only if } \{j | f(j) = g(j)\} \in \mathscr{F}$$

If we think of the sets in $\mathscr{F}$ as being 'large' sets, then, borrowing a phrase from measure theory, we read $f =_{\mathscr{F}} g$ as '$f(j) = g(j)$ almost everywhere'.

It is easy to see that $=_{\mathscr{F}}$ is an equivalence relation: (1) $f =_{\mathscr{F}} f$; (2) if $f =_{\mathscr{F}} g$ then $g =_{\mathscr{F}} f$; (3) if $f =_{\mathscr{F}} g$ and $g =_{\mathscr{F}} h$, then $f =_{\mathscr{F}} h$. For the proof of (3), observe that $\{j|f(j) = g(j)\} \cap \{j|g(j) = h(j)\} \subseteq \{j|f(j) = h(j)\}$. If $\{j|f(j) = g(j)\}$ and $\{j|g(j) = h(j)\}$ are in $\mathscr{F}$, then so is their intersection and, therefore, also $\{j|f(j) = h(j)\}$.

On the basis of the equivalence relation $=_{\mathscr{F}}$, we can divide $\Pi_{j\in J}D_j$ into equivalence classes: for any f in $\Pi_{j\in J}D_j$, we define its equivalence class $f_{\mathscr{F}}$ as $\{g|f =_{\mathscr{F}} g\}$. Clearly, (1) $f \in f_{\mathscr{F}}$; (2) $f_{\mathscr{F}} = h_{\mathscr{F}}$ if and only if $f =_{\mathscr{F}} h$; and (3) if $f_{\mathscr{F}} \neq h_{\mathscr{F}}$, then $f_{\mathscr{F}} \cap h_{\mathscr{F}} = \emptyset$. We denote the set of equivalence classes $f_{\mathscr{F}}$ by $\Pi_{j\in J}D_j/\mathscr{F}$. Intuitively, $\Pi_{j\in J}D_j/\mathscr{F}$ is obtained from $\Pi_{j\in J}D_j$ by identifying (or merging) elements of $\Pi_{j\in J}D_j$ that are equal almost everywhere.

Now we shall define a model M of K with domain $\Pi_{j\in J}D_j/\mathscr{F}$.

1. Let c be any individual constant of K and let c_j be the interpretation of c in M_j. Then the interpretation of c in M will be $f_{\mathscr{F}}$, where f is the function such that $f(j) = c_j$ for all j in J. We denote f by $\{c_j\}_{j\in J}$.
2. Let f_k^n be any function letter of K and let A_k^n be any predicate letter of K. Their interpretations $(f_k^n)^{\mathrm{M}}$ and $(A_k^n)^{\mathrm{M}}$ are defined in the following manner. Let $(g_1)_{\mathscr{F}}, \ldots, (g_n)_{\mathscr{F}}$ be any members of $\Pi_{j\in J}D_j/\mathscr{F}$.

(a) $(f_k^n)^{\mathrm{M}}((g_1)_{\mathscr{F}}, \ldots, (g_n)_{\mathscr{F}}) = h_{\mathscr{F}}$, where $h(j) = (f_k^n)^{\mathrm{M}_j}(g_1(j), \ldots, g_n(j))$ for all j in J.

(b) $(A_k^n)^{\mathrm{M}}((g_1)_{\mathscr{F}}, \ldots, (g_n)_{\mathscr{F}})$ holds if and only if $\{j| \models_{\mathrm{M}_j} A_k^n[g_1(j), \ldots, g_n(j)]\} \in \mathscr{F}$.

Intuitively, $(f_k^n)^{\mathrm{M}}$ is calculated componentwise, and $(A_k^n)^{\mathrm{M}}$ holds if and only if A_k^n holds in almost all components. Definitions (a) and (b) have to be shown to be independent of the choice of the representatives $g_1, \ldots, g_n$ in the equivalence classes $(g_1)_{\mathscr{F}}, \ldots, (g_n)_{\mathscr{F}}$: if $g_1 =_{\mathscr{F}} g_1^*, \ldots, g_n =_{\mathscr{F}} g_n^*$, and $h^*(j) = (f_k^n)^{\mathrm{M}_j}(g_1^*(j), \ldots, g_n^*(j))$, then (i) $h_{\mathscr{F}} =_{\mathscr{F}} h_{\mathscr{F}}^*$ and (ii) $\{j| \models_{\mathrm{M}_j} A_k^n[g_1(j), \ldots, g_n(j)]\} \in \mathscr{F}$ if and only if $\{j| \models_{\mathrm{M}_j} A_k^n[g_1^*(j), \ldots, g_n^*(j)]\} \in \mathscr{F}$.

Part (i) follows from the inclusion

$$\{j|g_1(j) = g_1^*(j)\} \cap \ldots \cap \{j|g_n(j) = g_n^*(j)\} \subseteq \{j|(f_k^n)^{\mathrm{M}_j}(g_1(j), \ldots, g_n(j)) = (f_k^n)^{\mathrm{M}_j}(g_1^*(j), \ldots, g_n^*(j))\}$$

Part (ii) follows from the inclusions:

$$\{j|g_1(j) = g_1^*(j)\} \cap \ldots \cap \{j|g_n(j) = g_n^*(j)\} \subseteq \{j| \models_{\mathrm{M}_j} A_k^n[g_1(j), \ldots, g_n(j)] \text{ if and only if } \models_{\mathrm{M}_j} A_k^n[g_1^*(j), \ldots, g_n^*(j)]\}$$

and

$$\{j| \models_{\mathrm{M}_j} A_k^n[g_1(j), \ldots, g_n(j)]\} \cap \{j| \models_{\mathrm{M}_j} A_k^n[g_1(j), \ldots, g_n(j)] \text{ if and only if } \models_{\mathrm{M}_j} A_k^n[g_1^*(j), \ldots, g_n^*(j)] \subseteq \{j| \models_{\mathrm{M}_j} A_k^n[g_1^*(j), \ldots, g_n^*(j)]\}$$

In the case of the equality relation $=$, which is an abbreviation for A_1^2,

$$(A_1^2)^{\mathrm{M}}(g_{\mathscr{F}}, h_{\mathscr{F}}) \text{ if and only if } \{j| \models_{\mathrm{M}_j} A_1^2[g(j), h(j)]\} \in \mathscr{F}$$
$$\text{if and only if } \{j|g(j) = h(j)\} \in \mathscr{F}$$
$$\text{if and only if } g =_{\mathscr{F}} h$$

that is, if and only if $g_{\mathscr{F}} = h_{\mathscr{F}}$. Hence, the interpretation $(A_1^2)^{\mathrm{M}}$ is the identity relation and the model M is normal.

The model M just defined will be denoted $\Pi_{j\in J}\mathrm{M}_j/\mathscr{F}$ and will be called a *reduced direct product*. When $\mathscr{F}$ is an ultrafilter, $\Pi_{j\in J}\mathrm{M}_j/\mathscr{F}$ is called an *ultraproduct*. When $\mathscr{F}$ is an ultrafilter and all the M_js are the same model N, then $\Pi_{j\in J}\mathrm{M}_j/\mathscr{F}$ is denoted $\mathrm{N}^J/\mathscr{F}$ and is called an *ultrapower*.

Examples

1. Choose a fixed element r of the index set J, and let $\mathscr{F}$ be the principal ultrafilter $\mathscr{F}_r = \{B|r \in B \wedge B \subseteq J\}$. Then, for any f, g in $\Pi_{j\in J}D_j, f =_{\mathscr{F}} g$ if and only if $\{j|f(j) = g(j)\} \in \mathscr{F}$, that is, if and only if $f(r) = g(r)$. Hence, a member of $\Pi_{j\in J}D_j/\mathscr{F}$ consists of all f in $\Pi_{j\in J}D_j$ that have the same rth component. For any predicate letter A_k^n of K and any $g_1, \ldots, g_n$ in $\Pi_{j\in J}D_j$, $\models_{\mathrm{M}} A_k^n[(g_1)_{\mathscr{F}}, \ldots, (g_n)_{\mathscr{F}}]$ if and only if $\{j| \models_{\mathrm{M}_j} A_k^n[g_1(j), \ldots, g_n(j)]\} \in \mathscr{F}$, that is, if and only if $\models_{\mathrm{M}_r} A_k^n[g_1(j), \ldots, g_n(j)]$. Hence, it is easy to verify that the function $\varphi : \Pi_{j\in J}D_j/\mathscr{F} \to D_r$, defined by $\varphi(g_{\mathscr{F}}) = g(r)$ is an isomorphism of $\Pi_{j\in J}\mathrm{M}_j/\mathscr{F}$ with M_r. Thus, when $\mathscr{F}$ is a principal ultrafilter, the ultraproduct $\Pi_{j\in J}\mathrm{M}_j/\mathscr{F}$ is essentially the same as one of its components and yields nothing new.
2. Let $\mathscr{F}$ be the filter $\{J\}$. Then, for any f, g in $\Pi_{j\in J}D_j, f =_{\mathscr{F}} g$ if and only if $\{j|f(j) = g(j)\} \in \mathscr{F}$, that is, if and only if $f(j) = g(j)$ for all j in J, or if and only if $f = g$. Thus, every member of $\Pi_{j\in J}D_j/\mathscr{F}$ is a singleton $\{g\}$ for some g in $\Pi_{j\in J}D_j$. Moreover, $(f_k^n)^M((g_1)_{\mathscr{F}}, \ldots, (g_n)_{\mathscr{F}}) = \{g\}$, where g is such that $g(j) = (f_k^n)^{\mathrm{M}_j}(g_1(j), \ldots, g_n(j))$ for all j in J. Also, $\models_{\mathrm{M}} A_k^n[(g_1)_{\mathscr{F}}, \ldots, (g_n)_{\mathscr{F}}]$ if and only if $\models_{\mathrm{M}_j} A_k^n[g_1(j), \ldots, g_n(j)]$ for all j in J. Hence, $\Pi_{j\in J}\mathrm{M}_j/_{\mathscr{F}}$ is, in this case, essentially the same as the ordinary 'direct product' $\Pi_{j\in J}\mathrm{M}_j$, in which the operations and relations are defined componentwise.
3. Let $\mathscr{F}$ be the improper filter $\mathscr{P}(J)$. Then, for any f, g in $\Pi_{j\in J}D_j, f =_{\mathscr{F}} g$ if and only if $\{j|f(j) = g(j)\} \in \mathscr{F}$, that is, if and only if $\{j|f(j) = g(j)\} \in \mathscr{P}(J)$. Thus, $f =_{\mathscr{F}} g$ for all f and g, and $\Pi_{j\in J}D_j/\mathscr{F}$ consists of only one element. For any predicate letter $A_k^n, \models_{\mathrm{M}} A_k^n[f_{\mathscr{F}}, \ldots, f_{\mathscr{F}}]$ if and only if $\{j \models_{\mathrm{M}_j} A_k^n[f(j), \ldots, f(j)])\} \in \mathscr{P}(J)$; that is, every atomic wf is true.

The basic theorem on ultraproducts is due to Loś (1955b).

PROPOSITION 2.42 (LOŚ'S THEOREM)

Let $\mathscr{F}$ be an ultrafilter on a set J and let $\mathrm{M} = \Pi_{j\in J}\mathrm{M}_j/\mathscr{F}$ be an ultraproduct.

(a) Let $s = ((g_1)_{\mathscr{F}}, (g_2)_{\mathscr{F}}, \ldots)$ be a denumerable sequence of elements of $\Pi_{j\in J} D_j/\mathscr{F}$. For each j in J, let s_j be the denumerable sequence $(g_1(j), g_2(j), \ldots)$ in D_j. Then, for any wf $\mathscr{B}$ of K, s satisfies $\mathscr{B}$ in M if and only if $\{j | s_j$ satisfies $\mathscr{B}$ in $\mathrm{M}_j\} \in \mathscr{F}$.

(b) For any sentence $\mathscr{B}$ of K, $\mathscr{B}$ is true in $\Pi_{j\in J}\mathrm{M}_j/\mathscr{F}$ if and only if $\{j | \models_{\mathrm{M}_j} \mathscr{B}\} \in \mathscr{F}$. (Thus, (b) asserts that a sentence $\mathscr{B}$ is true in an ultraproduct if and only if it is true in almost all components.)

Proof

(a) We shall use induction on the number m of connectives and quantifiers in $\mathscr{B}$. We can reduce the case $m = 0$ to the following subcases:† (i) $A_k^n(x_{i_1}, \ldots, x_{i_n})$; (ii) $x_\ell = f_k^n(x_{i_1}, \ldots, x_{i_n})$; and (iii) $x_\ell = a_k$. For subcase (i), s satisfies $A_k^n(x_{i_1}, \ldots, x_{i_n})$ if and only if $\models_{\mathrm{M}} A_k^n[(g_{i_1})_{\mathscr{F}}, \ldots, (g_{i_n})_{\mathscr{F}}]$, which is equivalent to $\{j | \models_{\mathrm{M}_j} A_k^n[g_{i_1}(j), \ldots, g_{i_n}(j)]\} \in \mathscr{F}$; that is $\{j | s_j$ satisfies $A_k^n(x_{i_1}, \ldots, x_{i_n})$ in $\mathrm{M}_j\} \in \mathscr{F}$. Subcases (ii) and (iii) are handled in similar fashion.

Now, let us assume the result holds for all wfs that have fewer than m connectives and quantifiers.

Case 1. $\mathscr{B}$ is $\neg\mathscr{C}$. By inductive hypothesis, s satisfies $\mathscr{C}$ in M if and only if $\{j | s_j$ satisfies $\mathscr{C}$ in $\mathrm{M}_j\} \in \mathscr{F}$. s satisfies $\neg\mathscr{C}$ in M if and only if $\{j | s_j$ satisfies $\mathscr{C}$ in $\mathrm{M}_j\} \notin \mathscr{F}$. But, since $\mathscr{F}$ is an ultrafilter, the last condition is equivalent, by exercise 2.119, to $\{j | s_j$ satisfies $\neg\mathscr{C}$ in $\mathrm{M}_j\} \in \mathscr{F}$.

Case 2. $\mathscr{B}$ is $\mathscr{C} \wedge \mathscr{D}$. By inductive hypothesis, s satisfies $\mathscr{C}$ in M if and only if $\{j | s_j$ satisfies $\mathscr{C}$ in $M_j\} \in \mathscr{F}$, and s satisfies $\mathscr{D}$ in M if and only if $\{j | s_j$ satisfies $\mathscr{D}$ in $\mathrm{M}_j\} \in \mathscr{F}$. Therefore, s satisfies $\mathscr{C} \wedge \mathscr{D}$ if and only if both of the indicated sets belong to $\mathscr{F}$. But, this is equivalent to their intersection belonging to $\mathscr{F}$, which, in turn, is equivalent to $\{j | s_j$ satisfies $\mathscr{C} \wedge \mathscr{D}$ in $\mathrm{M}_j\} \in \mathscr{F}$.

Case 3. $\mathscr{B}$ is $(\exists x_i)\mathscr{C}$. Assume s satisfies $(\exists x_i)\mathscr{C}$. Then there exists h in $\Pi_{j\in J} D_j$ such that s' satisfies $\mathscr{C}$ in M, where s' is the same as s except that $h_{\mathscr{F}}$ is the ith component of s'. By inductive hypothesis, s' satisfies $\mathscr{C}$ in M if and only if $\{j | s'_j$ satisfies $\mathscr{C}$ in $\mathrm{M}_j\} \in \mathscr{F}$. Hence, $\{j | s_j$ satisfies $(\exists x_i)\mathscr{C}$ in $\mathrm{M}_j\} \in \mathscr{F}$, since, if s'_j satisfies $\mathscr{C}$ in M_j then s_j satisfies $(\exists x_i)\mathscr{C}$ in M_j.

Conversely, assume $W = \{j | s_j$ satisfies$(\exists x_i)\mathscr{C}$ in $\mathrm{M}_j\} \in \mathscr{F}$. For each j in W, choose some s'_j such that s'_j is the same as s_j except in at most the ith component and s'_j satisfies $\mathscr{C}$. Now define h in $\Pi_{j\in J} D_j$ as follows: for j in W, let $h(j)$ be the ith component of s'_j, and , for $j \notin W$, choose $h(j)$ to be an

†A wf $A_k^n(t_1, \ldots, t_n)$ can be replaced by $(\forall u_1)\ldots(\forall u_n)(u_1 = t_1 \wedge \ldots \wedge u_n = t_n \Rightarrow A_k^n(u_1, \ldots, u_n))$, and a wf $x = f_k^n(t_1, \ldots, t_n)$ can be replaced by $(\forall z_1)\ldots(\forall z_n)(z_1 = t_1 \wedge \ldots \wedge z_n = t_n \Rightarrow x = f_k^n(z_1, \ldots, z_n))$. In this way, every wf is equivalent to a wf built up from wfs of the forms (i)–(iii) by applying connectives and quantifiers.

arbitrary element of D_j. Let s'' be the same as s except that its ith component is $h_{\mathscr{F}}$. Then $W \subseteq \{j | s''_j$ satisfies $\mathscr{C}$ in $\mathrm{M}_j\} \in \mathscr{F}$. Hence, by the inductive hypothesis, s'' satisfies $\mathscr{C}$ in M. Therefore, s satisfies $(\exists x_i)\mathscr{C}$ in M.

(b) This follows from part (a) by noting that a sentence $\mathscr{B}$ is true in a model if and only if some sequence satisfies $\mathscr{B}$.

COROLLARY 2.43

If M is a model and $\mathscr{F}$ is an ultrafilter on J, and if M^* is the ultrapower $\mathrm{M}^J/\mathscr{F}$, then $\mathrm{M}^* \equiv \mathrm{M}$.

Proof

Let $\mathscr{B}$ be any sentence. Then, by Proposition 2.42(b), $\mathscr{B}$ is true in M^* if and only if $\{j | \mathscr{B}$ is true in M $\} \in \mathscr{F}$. If $\mathscr{B}$ is true in M, $\{j | \mathscr{B}$ is true in M $\} = J \in \mathscr{F}$. If $\mathscr{B}$ is false in M, $\{j | \mathscr{B}$ is true in M $\} = \emptyset \notin \mathscr{F}$.

Corollary 2.43 can be strengthened considerably. For each c in the domain D of M, let $c^{\#}$ stand for the constant function such that $c^{\#}(j) = c$ for all j in J. Define the function ψ such that, for each c in D, $\psi(c) = (c^{\#})_{\mathscr{F}} \in D^J/\mathscr{F}$, and denote the range of ψ by $\mathrm{M}^{\#}$. $\mathrm{M}^{\#}$ obviously contains the interpretations in M^* of the individual constants. Moreover, $\mathrm{M}^{\#}$ is closed under the operations $(f_k^n)^{\mathrm{M}^*}$; for $(f_k^n)^{\mathrm{M}^*}((c_1^{\#})_{\mathscr{F}}, \ldots, (c_n^{\#})_{\mathscr{F}})$ is $h_{\mathscr{F}}$, where $h(j) = (f_k^n)^{\mathrm{M}}(c_1, \ldots, c_n)$ for all j in J, and $(f_k^n)^{\mathrm{M}}((c_1, \ldots, c_n)$ is a fixed element b of D. So, $h_{\mathscr{F}} = (b^{\#})_{\mathscr{F}} \in \mathrm{M}^{\#}$. Thus, $\mathrm{M}^{\#}$ is a substructure of M^*.

COROLLARY 2.44

ψ is an isomorphism of M with $\mathrm{M}^{\#}$, and $\mathrm{M}^{\#} \leqslant_{\mathrm{e}} \mathrm{M}^*$.

Proof

(a) By definition of $\mathrm{M}^{\#}$, the range of ψ is $\mathrm{M}^{\#}$.

(b) ψ is one–one. (For any c, d in D, $(c^{\#})_{\mathscr{F}} = (d^{\#})_{\mathscr{F}}$ if and only if $c^{\#} =_{\mathscr{F}} d^{\#}$, which is equivalent to $\{j | c^{\#}(j) = d^{\#}(j)\} \in \mathscr{F}$; that is, $\{j | c = d\} \in \mathscr{F}$. If $c \neq d$, $\{j | c = d\} = \emptyset \notin \mathscr{F}$, and, therefore, $\psi(c) \neq \psi(d)$.

(c) For any $c_1, \ldots, c_n$ in D, $(f_k^n)^{\mathrm{M}^*}(\psi(c_1) \ldots, \psi(c_n)) = (f_k^n)^{\mathrm{M}^*}((c_1^{\#})_{\mathscr{F}}, \ldots, (c_n^{\#})_{\mathscr{F}}) = h_{\mathscr{F}}$, where $h(j) = (f_k^n)^{\mathrm{M}}(c_1^{\#}(j), \ldots, c_n^{\#}(j)) = (f_k^n)^{\mathrm{M}}(c_1, \ldots, c_n)$. Thus, $h_{\mathscr{F}} = ((f_k^n)^{\mathrm{M}}(c_1, \ldots, c_n))^{\#}/\mathscr{F} = \psi((f_k^n)^{\mathrm{M}}(c_1, \ldots, c_n))$.

(d) $\models_{\mathrm{M}^*} A_k^n[\psi(c_1) \ldots, \psi(c_n)]$ if and only if $\{j | \models_{\mathrm{M}} A_k^n(\psi(c_1)(j), \ldots, \psi(c_n)(j))\} \in \mathscr{F}$, which is equivalent to $\{j | \models_{\mathrm{M}} A_k^n(c_1, \ldots, c_n)\} \in \mathscr{F}$, that is, $\models_{\mathrm{M}} A_k^n[c_1, \ldots, c_n]$. Thus, ψ is an isomorphism of M with $\mathrm{M}^{\#}$.

To see that $M^{\#} \leqslant_e M^*$, let $\mathscr{B}$ be any wf and $(c_1^{\#})_{\mathscr{F}}, \ldots, (c_n^{\#})_{\mathscr{F}} \in M^{\#}$. Then, by proposition 2.42(a), $\models_{M^*} \mathscr{B}[(c_1^{\#})_{\mathscr{F}}, \ldots (c_n^{\#})_{\mathscr{F}}]$ if and only if $\{j| \models_M \mathscr{B}[c_1^{\#}(j), \ldots, c_n^{\#}(j)]\} \in \mathscr{F}$, which is equivalent to $\{j| \models_M \mathscr{B}[c_1, \ldots, c_n]\} \in \mathscr{F}$, which, in turn, is equivalent to $\models_M \mathscr{B}[c_1, \ldots, c_n]$, that is, to $\models_{M^{\#}} \mathscr{B}[(c_1^{\#})_{\mathscr{F}}, \ldots, (c_n^{\#})_{\mathscr{F}}]$, since ψ is an isomorphism of M with $M^{\#}$.

Exercises

2.124 (The compactness theorem again; see Exercise 2.54) If all finite subsets of a set of sentences Γ have a model, then Γ has a model.

2.125

(a) A class $\mathscr{W}$ of interpretations of a language $\mathscr{L}$ is called *elementary* if there is a set Γ of sentences of $\mathscr{L}$ such that $\mathscr{W}$ is the class of all models of Γ. Prove that $\mathscr{W}$ is elementary if and only if $\mathscr{W}$ is closed under elementary equivalence and the formation of ultraproducts.

(b) A class $\mathscr{W}$ of interpretations of a language $\mathscr{L}$ will be called *sentential* if there is a sentence $\mathscr{B}$ of $\mathscr{L}$ such that $\mathscr{W}$ is the class of all models of $\mathscr{B}$. Prove that a class $\mathscr{W}$ is sentential if and only if both $\mathscr{W}$ and its complement $\overline{\mathscr{W}}$ (all interpretations of $\mathscr{L}$ not in $\mathscr{W}$) are closed with respect to elementary equivalence and ultraproducts.

(c) Prove that theory K of fields of characteristic 0 (see page 117) is axiomatizable but not finitely axiomatizable.

Non-standard analysis

From the invention of the calculus until relatively recent times the idea of *infinitesimals* has been an intuitively meaningful tool for finding new results in analysis. The fact that there was no rigourous foundation for infinitesimals was a source of embarrassment and led mathematicians to discard them in favour of the rigorous limit ideas of Cauchy and Weierstrass. However, about forty years ago, Abraham Robinson discovered that it was possible to resurrect infinitesimals in an entirely legitimate and precise way. This can be done by constructing models that are elementarily equivalent to, but not isomorphic to, the ordered field of real numbers. Such models can be produced either by using Proposition 2.33 or as ultrapowers. We shall sketch here the method based on ultrapowers.

Let R be the set of real numbers. Let K be a generalized predicate calculus with equality having the following symbols:

1. For each real number r, there is an individual constant a_r.
2. For every n-ary operation φ on R, there is a function letter f_φ.
3. For every n-ary relation Φ on R, there is a predicate letter A_Φ.

We can think of R as forming the domain of a model $\mathscr{R}$ for K; we simply let $(a_r)^{\mathscr{R}} = r$, $(f_\varphi)^{\mathscr{R}} = \varphi$, and $(A_\Phi)^{\mathscr{R}} = \Phi$.

Let $\mathscr{F}$ be a non-principal ultrafilter on the set ω of natural numbers. We can then form the ultrapower $\mathscr{R}^* = \mathscr{R}^\omega/\mathscr{F}$. We denote the domain $R^\omega/\mathscr{F}$ of $\mathscr{R}^*$ by R^*. By Corollary 2.43, $\mathscr{R}^* \equiv \mathscr{R}$ and, therefore, $\mathscr{R}^*$ has all the properties formalizable in K that $\mathscr{R}$ possesses. Moreover, by Corollary 2.44, $\mathscr{R}^*$ has an elementary submodel $\mathscr{R}^\#$ that is an isomorphic image of $\mathscr{R}$. The domain $R^\#$ of $\mathscr{R}^\#$ consists of all elements $(c^\#)_\mathscr{F}$ corresponding to the constant functions $c^\#(i) = c$ for all i in ω. We shall sometimes refer to the members of $R^\#$ also as real numbers; the elements of $R^* - R^\#$ will be called *non-standard reals*.

That there exist non-standard reals can be shown by explicitly exhibiting one. Let $\imath(j) = j$ for all j in ω. Then $\imath_\mathscr{F} \in R^*$. However, $(c^\#)_\mathscr{F} < \imath_\mathscr{F}$ for all c in R, by virtue of Łoś's theorem and the fact that $\{j|c^\#(j) < \imath(j)\} = \{j|c < j\}$, being the set of all natural numbers greater than a fixed real number, is the complement of a finite set and is, therefore, in the non-principal ultrafilter $\mathscr{F}$. $\imath_\mathscr{F}$ is an 'infinitely large' non-standard real. (The relation $<$ used in the assertion $(c^\#)_\mathscr{F} < \imath_\mathscr{F}$ is the relation on the ultrapower $\mathscr{R}^*$ corresponding to the predicate letter $<$ of K. We use the symbol $<$ instead of $(<)^{\mathscr{R}^*}$ in order to avoid excessive notation, and we shall often do the same with other relations and functions, such as $u + v$, $u \times v$, and $|u|$.)

Since $\mathscr{R}^*$ possesses all the properties of $\mathscr{R}$ formalizable in K, $\mathscr{R}^*$ is an ordered field having the real number field $\mathscr{R}^\#$ as a proper subfield. ($\mathscr{R}^*$ is non-Archimedean: the element $\imath_\mathscr{F}$ defined above is greater than all the natural numbers $(n^\#)_\mathscr{F}$ of $\mathscr{R}^*$.) Let R_1, the set of 'finite' elements of R^*, contain those elements z such that $|z| < u$ for some real number u in $R^\#$. (R_1 is easily seen to form a subring of R^*.) Let R_0 consist of 0 and the 'infinitesimals' of R^*, that is, those elements $z \neq 0$ such that $|z| < u$ for all positive real numbers u in $R^\#$. The reciprocal $1/\imath_\mathscr{F}$ is an infinitesimal.) It is not difficult to verify that R_0 is an ideal in the ring R_1. In fact, since $x \in R_1 - R_0$ implies that $1/x \in R_1 - R_0$, it can be easily proved that R_0 is a maximal ideal in R_1.

Exercises

2.126 Prove that the cardinality of R^* is $2^{\aleph_0}$.

2.127 Prove that the set R_0 is closed under the operations of $+$, $-$ and $\times$.

2.128 Prove that, if $x \in R_1$ and $y \in R_0$, then $xy \in R_0$.

2.129 Prove that, if $x \in R_1 - R_0$, then $1/x \in R_1 - R_0$.

Let $x \in R_1$. Let $A = \{u|u \in R^\# \wedge u < x\}$ and $B = \{u|u \in R^\# \wedge u > x\}$. Then $\langle A, B\rangle$ is a 'cut' and, therefore, determines a unique real number r such that (1) $(\forall x)(x \in A \Rightarrow x \leqslant r)$ and (2) $(\forall x)(x \in B \Rightarrow x \geqslant r)$.† The difference $x - r$ is 0

†See Mendelson (1973, chap. 5).

or an infinitesimal. (Proof: Assume $x - r$ is not 0 or an infinitesimal. Then $|x - r| > r_1$ for some positive real number r_1. If $x > r$, then $x - r > r_1$. So $x > r + r_1 > r$. But then $r + r_1 \in A$, contradicting condition (1). If $x < r$, then $r - x > r_1$, and so $r > r - r_1 > x$. Thus, $r - r_1 \in B$, contradicting condition (2).) The real number r such that $x - r$ is 0 or an infinitesimal is called the *standard part* of x and is denoted $\mathrm{st}(x)$. Note that, if x is itself a real number, then $\mathrm{st}(x) = x$. We shall use the notation $x \approx y$ to mean $\mathrm{st}(x) = \mathrm{st}(y)$. Clearly, $x \approx y$ if and only if $x - y$ is 0 or an infinitesimal. If $x \approx y$, we say that x and y are *infinitely close*.

Exercises

2.130 If $x \in \mathrm{R}_1$, show that there is a unique real number r such that $x - r$ is 0 or an infinitesimal. (It is necessary to check this to ensure that $\mathrm{st}(x)$ is well-defined.)

2.131 If x and y are in R_1, prove the following.

(a) $\mathrm{st}(x + y) = \mathrm{st}(x) + \mathrm{st}(y)$
(b) $\mathrm{st}(xy) = \mathrm{st}(x)\mathrm{st}(y)$
(c) $\mathrm{st}(-x) = -\mathrm{st}(x) \wedge \mathrm{st}(y - x) = \mathrm{st}(y) - \mathrm{st}(x)$
(d) $x \geqslant 0 \Rightarrow \mathrm{st}(x) \geqslant 0$
(e) $x \leqslant y \Rightarrow \mathrm{st}(x) \leqslant \mathrm{st}(y)$

The set of natural numbers is a subset of the real numbers. Therefore, in the theory K there is a predicate letter N corresponding to the property $x \in \omega$. Hence, in R^*, there is a set ω^* of elements satisfying the wf $N(x)$. An element $f_{\mathscr{F}}$ of R^* satisfies $N(x)$ if and only if $\{j | f(j) \in \omega\} \in \mathscr{F}$. In particular, the elements $n^{\#}_{\mathscr{F}}$, for $n \in \omega$, are the 'standard' members of ω^*, whereas $\iota_{\mathscr{F}}$, for example, is a 'non-standard' natural number in R^*.

Many of the properties of the real number system can be studied from the viewpoint of non-standard analysis. For example, if s is an ordinary denumerable sequence of real numbers and c is a real number, one ordinarily says that $\lim s_n = c$ if

$$(\&) \quad (\forall \varepsilon)(\varepsilon > 0 \Rightarrow (\exists n)(n \in \omega \wedge (\forall k)(k \in \omega \wedge k \geqslant n \Rightarrow |s_k - c| < \varepsilon)))$$

Since $s \in R^{\omega}$, s is a relation and, therefore, the theory K contains a predicate letter $S(n, x)$ corresponding to the relation $s_n = x$. Hence, R^* will have a relation of all pairs $\langle n, x \rangle$ satisfying $S(n, x)$. Since $\mathscr{R}^* \equiv \mathscr{R}$, this relation will be a function that is an extension of the given sequence to the larger domain ω^*. Then we have the following result.

PROPOSITION 2.45

Let s be a denumerable sequence of real numbers and c a real number. Let s^* denote the function from ω^* into R^* corresponding to s in $\mathscr{R}^*$. Then $\lim s_n = c$ if and only if $s^*(n) \approx c$ for all n in $\omega^* - \omega$. (The latter condition can be paraphrased by saying that $s^*(n)$ is infinitely close to c when n is infinitely large.)

Proof

Assume $\lim s_n = c$. Consider any positive real ε. By (&), there is a natural number n_0 such that $(\forall k)(k \in \omega \wedge k \geqslant n_0 \Rightarrow |s_k - c| < \varepsilon)$ holds in $\mathscr{R}$. Hence, the corresponding sentence $(\forall k)(k \in \omega^* \wedge k \geqslant n_0 \Rightarrow |s^*(k) - c| < \varepsilon)$ holds in $\mathscr{R}^*$. For any n in $\omega^* - \omega$, $n > n_0$ and, therefore, $|s^*(n) - c| < \varepsilon$. Since this holds for all positive real ε, $s^*(n) - c$ is 0 or an infinitesimal.

Conversely, assume $s^*(n) \approx c$ for all $n \in \omega^* - \omega$. Take any positive real ε. Fix some n_1 in $\omega^* - \omega$. Then $(\forall k)(k \geqslant n_1 \Rightarrow |s^*(k) - c| < \varepsilon)$. So the sentence $(\exists n)(n \in \omega \wedge (\forall k)(k \in \omega \wedge k \geqslant n \Rightarrow |s_k - c| < \varepsilon))$ is true for $\mathscr{R}^*$ and, therefore, also for $\mathscr{R}$. So there must be a natural number n_0 such that $(\forall k)(k \in \omega \wedge k \geqslant n_0 \Rightarrow |s_k - c| < \varepsilon)$. Since ε was an arbitrary positive real number, we have proved $\lim s_n = c$.

Exercise

2.132 Using Proposition 2.45, prove the following limit theorems for the real number system. If s and u are denumerable sequences of real numbers and c_1 and c_2 are real numbers such that $\lim s_n = c_1$ and $\lim u_n = c_2$, then:

(a) $\lim(s_n + u_n) = c_1 + c_2$;
(b) $\lim(s_n u_n) = c_1 c_2$;
(c) If $c_2 \neq 0$ and all $u_n \neq 0$, $\lim(s_n/u_n) = c_1/c_2$.

Let us now consider another important notion of analysis, continuity. Let B be a set of real numbers, let $c \in B$, and let f be a function defined on B and taking real values. One says that f is *continuous* at c if

$$(\P) \quad (\forall \varepsilon)(\varepsilon > 0 \Rightarrow (\exists \delta)(\delta > 0 \wedge (\forall x)(x \in B \wedge |x - c| < \delta \Rightarrow |\mathrm{f}(x) - \mathrm{f}(c)| < \varepsilon)))$$

PROPOSITION 2.46

Let f be a real-valued function on a set B of real numbers. Let $c \in B$. Let B^* be the subset of R^* corresponding to B, and let f^* be the function corresponding to f.† Then f is continuous at c if and only if $(\forall x)(x \in B^* \wedge x \approx c \Rightarrow f^*(x) \approx f(c))$.

Exercises

2.133 Prove Proposition 2.46.

2.134 Assume f and g are real-valued functions defined on a set B of real numbers and assume that f and g are continuous at a point c in B. Using Proposition 2.46, prove the following.

(a) $f + g$ is continuous at c.
(b) $f \cdot g$ is continuous at c.

2.135 Let f be a real-valued function defined on a set B of real numbers and continuous at a point c in B, and let g be a real-valued function defined on a set A of real numbers containing the image of B under f. Assume that g is continuous at the point $f(c)$. Prove, by Proposition 2.46, that the composition $g \circ f$ is continuous at c.

2.136 Let $C \subseteq R$.

(a) C is said to be *closed* if $(\forall x)((\forall \varepsilon)[\varepsilon > 0 \Rightarrow (\exists y)(y \in C \wedge |x - y| < \varepsilon)] \Rightarrow x \in C)$. Show that C is closed if and only if every real number that is infinitely close to a member of C^* is in C. (b) C is said to be *open* if $(\forall x)(x \in C \Rightarrow (\exists \delta)(\delta > 0 \wedge (\forall y)(|y - x| < \delta \Rightarrow y \in C)))$. Show that C is open if and only if every non-standard real number that is infinitely close to a member of C is a member of C^*.

Many standard theorems of analysis turn out to have much simpler proofs within non-standard analysis. Even stronger results can be obtained by starting with a theory K that has symbols, not only for the elements, operations and relations on R, but also for sets of subsets of R, sets of sets of subsets of R, and so on. In this way, the methods of non-standard analysis can be applied to all areas of modern analysis, sometimes with original and striking results. For further development and applications, see A. Robinson (1966), Luxemburg (1969), Bernstein (1973), Stroyan and Luxemburg (1976), and Davis (1977a). A calculus textbook based on non-standard analysis has been written by Keisler (1976) and has been used in some experimental undergraduate courses.

†To be more precise, f is represented in the theory K by a predicate letter A_f, where $A_f(x, y)$ corresponds to the relation $f(x) = y$. Then the corresponding relation A_f^* in R^* determines a function f^* with domain B^*.

Exercises

2.137 A real-valued function f defined on a closed interval $[a,b] = \{x|a \leqslant x \leqslant b\}$ is said to be *uniformly continuous* if

$$(\forall\varepsilon)(\varepsilon > 0 \Rightarrow (\exists\delta)(\delta > 0 \wedge (\forall x)(\forall y)(a \leqslant x \leqslant b \wedge a \leqslant y \leqslant b \wedge |x-y| < \delta \Rightarrow |f(x)-f(y)| < \varepsilon)))$$

Prove that f is uniformly continuous if and only if, for all x and y in $[a,b]^*$, $x \approx y \Rightarrow f^*(x) \approx f^*(y)$.

2.138 Prove by non-standard methods that any function continuous on $[a,b]$ is uniformly continuous on $[a,b]$.

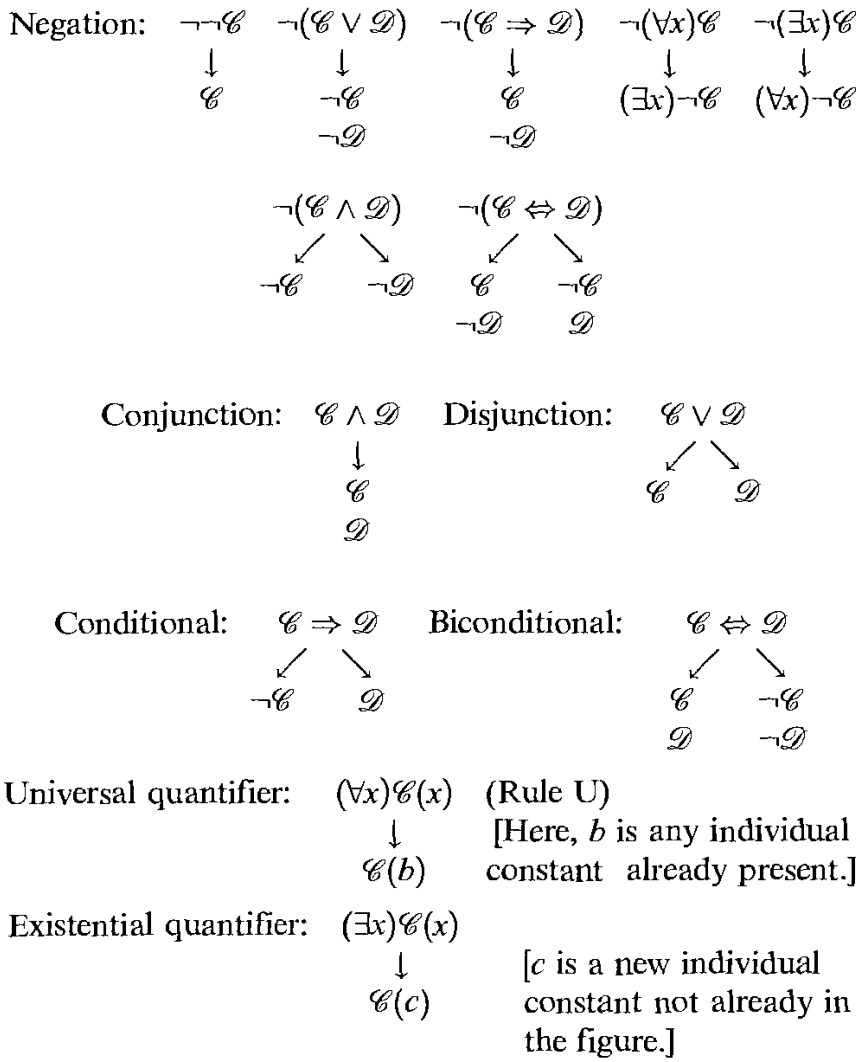
2.15 SEMANTIC TREES

Remember that a wf is logically valid if and only if it is true for all interpretations. Since there are uncountably many interpretations, there is no simple direct way to determine logical validity. Gödel's completeness theorem (Corollary 2.19) showed that logical validity is equivalent to derivability in a predicate calculus. But, to find out whether a wf is provable in a predicate calculus, we have only a very clumsy method: start generating the theorems and watch to see whether the given wf ever appears. Our aim here is to outline a more intuitive and usable approach in the case of wfs without function letters. Throughout this section, we assume that no function letters occur in our wfs.

A wf is logically valid if and only if its negation is not satisfiable. We shall now explain a simple procedure for trying to determine satisfiability of a closed wf $\mathscr{B}$.† Our purpose is either to show that $\mathscr{B}$ is not satisfiable or to find a model for $\mathscr{B}$.

We shall construct a figure in the shape of an inverted tree. Start with the wf $\mathscr{B}$ at the top (the 'root' of the tree). We apply certain rules for writing wfs below those already obtained. These rules replace complicated wfs by simpler ones in a way that corresponds to the meaning of the connectives and quantifiers.

†Remember that a wf is logically valid if and only if its closure is logically valid. So it suffices to consider only closed wfs.



Note that some of the rules require a fork or branching. This occurs when the given wf implies that one of two possible situations holds.

A *branch* is a sequence of wfs starting at the top and proceeding down the figure by applications of the rules. When a wf and its negation appear in a branch, that branch becomes *closed* and no further rules need be applied to the wf at the end of the branch. Closure of a branch will be indicated by a large cross $\times$.

Inspection of the rules shows that, when a rule is applied to a wf, the usefulness of that wf has been exhausted (the formula will be said to be *discharged*) and that formula need never be subject to a rule again, except in the case of a universally quantified wf. In the latter case, whenever a new individual constant appears in a branch below the wf, rule U can be applied with that new constant. *In addition, if no further rule applications are possible along a branch and no individual constant occurs in that branch, then we must introduce a new individual constant for use in possible applications of rule U along that branch.* (The idea behind this requirement is that, if we are trying to build a model, we must introduce a symbol for at least one object that can belong to the domain of the model.)

BASIC PRINCIPLE OF SEMANTIC TREES

If all branches become closed, the original wf is unsatisfiable. If, however, a branch remains unclosed, that branch can be used to construct a model in which the original wf is true; the domain of the model consists of the individual constants that appear in that branch.

We shall discuss the justification of this principle later on. First, we shall give examples of its use.

Examples

1. To prove that $(\forall x)\mathscr{C}(x) \Rightarrow \mathscr{C}(b)$ is logically valid, we build a semantic tree starting from its negation.

(i)	$\neg((\forall x)\mathscr{C}(x) \Rightarrow \mathscr{C}(b))$	
(ii)	$(\forall x)\mathscr{C}(x)$	(i)
(iii)	$\neg\mathscr{C}(b)$	(i)
(iv)	$\mathscr{C}(b)$	(ii)
	$\times$	

The number to the right of a given wf indicates the number of the line of the wf from which the given wf is derived. Since the only branch in this tree is closed, $\neg((\forall x)\mathscr{C}(x) \Rightarrow \mathscr{C}(b))$ is unsatisfiable and, therefore, $(\forall x)\mathscr{C}(x) \Rightarrow \mathscr{C}(b))$ is logically valid.

2.

(i)	$\neg[(\forall x)(\mathscr{C}(x) \Rightarrow \mathscr{D}(x)) \Rightarrow ((\forall x)\mathscr{C}(x) \Rightarrow (\forall x)\mathscr{D}(x))]$	
(ii)	$(\forall x)(\mathscr{C}(x) \Rightarrow \mathscr{D}(x))$	(i)
(iii)	$\neg((\forall x)\mathscr{C}(x) \Rightarrow (\forall x)\mathscr{D}(x))$	(i)
(iv)	$(\forall x)\mathscr{C}(x)$	(iii)
(v)	$\neg(\forall x)\mathscr{D}(x)$	(iii)
(vi)	$(\exists x)\neg\mathscr{D}(x)$	(v)
(vii)	$\neg\mathscr{D}(b)$	(vi)
(viii)	$\mathscr{C}(b)$	(iv)
(ix)	$\mathscr{C}(b) \Rightarrow \mathscr{D}(b)$	(ii)
	$\swarrow \quad \searrow$	
(x)	$\neg\mathscr{C}(b) \quad \mathscr{D}(b)$	(ix)
	$\times \quad \times$	

Since both branches are closed, the original wf (i) is unsatisfiable and, therefore, $(\forall x)(\mathscr{C}(x) \Rightarrow \mathscr{D}(x)) \Rightarrow ((\forall x)\mathscr{C}(x) \Rightarrow (\forall x)\mathscr{D}(x))$ is logically valid.

3.

(i)	$\neg[(\exists x)A_1^1(x) \Rightarrow (\forall x)A_1^1(x)]$	
(ii)	$(\exists x)A_1^1(x)$	(i)
(iii)	$\neg(\forall x)A_1^1(x)$	(i)
(iv)	$A_1^1(b)$	(ii)
(v)	$(\exists x)\neg A_1^1(x)$	(iii)
(vi)	$\neg A_1^1(c)$	(v)

No further applications of rules are possible and there is still an open branch. Define a model M with domain $\{b, c\}$ such that the interpretation of A_1^1 holds for b but not for c. Thus, $(\exists x)\neg A_1^1(x)$ is true in M but $(\forall x)A_1^1(x)$ is false in M. Hence, $(\exists x)A_1^1(x) \Rightarrow (\forall x)A_1^1(x)$ is false in M and is, therefore, not logically valid.

4. (i) $\neg[(\exists y)(\forall x)\mathscr{B}(x,y) \Rightarrow (\forall x)(\exists y)\mathscr{B}(x,y)]$
   (ii) $(\exists y)(\forall x)\mathscr{B}(x,y)$ (i)
   (iii) $\neg(\forall x)(\exists y)\mathscr{B}(x,y)$ (i)
   (iv) $(\forall x)\mathscr{B}(x,b)$ (ii)
   (v) $(\exists x)\neg(\exists y)\mathscr{B}(x,y)$ (iii)
   (vi) $\mathscr{B}(b,b)$ (iv)
   (vii) $\neg(\exists y)\mathscr{B}(c,y)$ (v)
   (viii) $\mathscr{B}(c,b)$ (iv)
   (ix) $(\forall y)\neg\mathscr{B}(c,y)$ (vii)
   (x) $\neg\mathscr{B}(c,b)$ (ix)
   $\times$

Hence, $(\exists y)(\forall x)\mathscr{B}(x,y) \Rightarrow (\forall x)(\exists y)\mathscr{B}(x,y)$ is logically valid.

Notice that, in the last tree, step (vi) served no purpose but was required by our method of constructing trees. We should be a little more precise in describing that method. At each step, we apply the appropriate rule to each undischarged wf (except universally quantified wfs), starting from the top of the tree. Then, to every universally quantified wf on a given branch we apply rule U with every individual constant that has appeared on that branch since the last step. In every application of a rule to a given wf, we write the resulting wf(s) below the branch that contains that wf.

5. (i) $\neg[(\forall x)\mathscr{B}(x) \Rightarrow (\exists x)\mathscr{B}(x)]$
   (ii) $(\forall x)\mathscr{B}(x)$ (i)
   (iii) $\neg(\exists x)\mathscr{B}(x)$ (i)
   (iv) $(\forall x)\neg\mathscr{B}(x)$ (iii)
   (v) $\mathscr{B}(b)$ (ii)†
   (vi) $\neg\mathscr{B}(b)$ (iv)
   $\times$

Hence, $(\forall x)\mathscr{B}(x) \Rightarrow (\exists x)\mathscr{B}(x)$ is logically valid.

6. (i) $\neg[(\forall x)\neg A_1^2(x,x) \Rightarrow (\exists x)(\forall y)\neg A_1^2(x,y)]$
   (ii) $(\forall x)\neg A_1^2(x,x)$ (i)
   (iii) $\neg(\exists x)(\forall y)\neg A_1^2(x,y)$ (ii)
   (iv) $(\forall x)\neg(\forall y)\neg A_1^2(x,y)$ (iii)
   (v) $\neg A_1^2(a_1,a_1)$ (ii)†
   (vi) $\neg(\forall y)\neg A_1^2(a_1,y)$ (iv)

†Here, we must introduce a new individual constant for use with rule U since, otherwise, the branch would end and would not contain any individual constants.

(vii)	$(\exists y)\neg\neg A_1^2(a_1, y)$	(vi)
(viii)	$\neg\neg A_1^2(a_1, a_2)$	(vii)
(ix)	$A_1^2(a_1, a_2)$	(viii)
(x)	$\neg A_1^2(a_2, a_2)$	(ii)
(xi)	$\neg(\forall y)\neg A_1^2(a_2, y)$	(iv)
(xii)	$(\exists y)\neg\neg A_1^2(a_2, y)$	(xi)
(xiii)	$\neg\neg A_1^2(a_2, a_3)$	(xii)
(xiv)	$A_1^2(a_2, a_3)$	(xiii)

We can see that the branch will never end and that we will obtain a sequence of constants $a_1, a_2, \ldots$ with wfs $A_1^2(a_n, a_{n+1})$ and $\neg A_1^2(a_n, a_n)$. Thus, we construct a model M with domain $\{a_1, a_2, \ldots\}$ and we define $(A_1^2)^{\mathrm{M}}$ to contain only the pairs $\langle a_n, a_{n+1}\rangle$. Then, $(\forall x)\neg A_1^2(x, x)$ is true in M, whereas $(\exists x)(\forall y)\neg A_1^2(x, y)$ is false in M. Hence, $(\forall x)\neg A_1^2(x, x) \Rightarrow (\exists x)(\forall y)\neg A_1^2(x, y)$ is not logically valid.

Exercises

2.139 Use semantic trees to determine whether the following wfs are logically valid.

(a) $(\forall x)(A_1^1(x) \vee A_2^1(x)) \Rightarrow ((\forall x)A_1^1(x)) \vee (\forall x)A_2^1(x)$
(b) $((\forall x)\mathscr{B}(x)) \wedge (\forall x)\mathscr{C}(x) \Rightarrow (\forall x)(\mathscr{B}(x) \wedge \mathscr{C}(x))$
(c) $(\forall x)(\mathscr{B}(x) \wedge \mathscr{C}(x)) \Rightarrow ((\forall x)\mathscr{B}(x)) \wedge (\forall x)\mathscr{C}(x)$
(d) $(\exists x)(A_1^1(x) \Rightarrow A_2^1(x)) \Rightarrow ((\exists x)A_1^1(x) \Rightarrow (\exists x)A_2^1(x))$
(e) $(\exists x)(\exists y)A_1^2(x, y) \Rightarrow (\exists z)A_1^2(z, z)$
(f) $((\forall x)A_1^1(x)) \vee (\forall x)A_2^1(x) \Rightarrow (\forall x)(A_1^1(x) \vee A_2^1(x))$
(g) $(\exists x)(\exists y)(A_1^2(x, y) \Rightarrow (\forall z)A_1^2(z, y))$
(h) The wfs of Exercises 2.24, 2.31(a, e, j), 2.39 and 2.40.
(i) The wfs of Exercise 2.21(a, b, g).

PROPOSITION 2.47

Assume that Γ is a set of closed wfs that satisfy the following closure conditions: (a) if $\neg\neg\mathscr{B}$ is in Γ, then $\mathscr{B}$ is in Γ; (b) if $\neg(\mathscr{B} \vee \mathscr{C})$ is in Γ, then $\neg\mathscr{B}$ and $\neg\mathscr{C}$ are in Γ; (c) if $\neg(\mathscr{B} \Rightarrow \mathscr{C})$ is in Γ, then $\mathscr{B}$ and $\neg\mathscr{C}$ are in Γ; (d) if $\neg(\forall x)\mathscr{B}$ is in Γ, then $(\exists x)\neg\mathscr{B}$ is in Γ; (e) if $\neg(\exists x)\mathscr{B}$ is in Γ, then $(\forall x)\neg\mathscr{B}$ is in Γ; (f) if $\neg(\mathscr{B} \wedge \mathscr{C})$ is in Γ, then at least one of $\neg\mathscr{B}$ and $\neg\mathscr{C}$ is in Γ; (g) if $\neg(\mathscr{B} \Leftrightarrow \mathscr{C})$ is in Γ, then either $\mathscr{B}$ and $\neg\mathscr{C}$ are in Γ, or $\neg\mathscr{B}$ and $\mathscr{C}$ are in Γ; (h) if $\mathscr{B} \wedge \mathscr{C}$ is in Γ, then so are $\mathscr{B}$ and $\mathscr{C}$; (i) if $\mathscr{B} \vee \mathscr{C}$ is in Γ, then at least one of $\mathscr{B}$ and $\mathscr{C}$ is in Γ, (j) if $\mathscr{B} \Rightarrow \mathscr{C}$ is in Γ, then at least one of $\neg\mathscr{B}$ and $\mathscr{C}$ is in Γ; (k) if $\mathscr{B} \Leftrightarrow \mathscr{C}$ is in Γ, then either $\mathscr{B}$ and $\mathscr{C}$ are in Γ or $\neg\mathscr{B}$ and $\neg\mathscr{C}$ are in Γ; (l) if $\forall x)\mathscr{B}(x)$ is in Γ, then $\mathscr{B}(b)$ is in Γ (where b is any individual constant that occurs in some wf of Γ); (m) if $(\exists x)\mathscr{B}(x)$ is in Γ, then $\mathscr{B}(b)$ is in Γ for some

individual constant b. If no wf and its negation both belong to Γ and some wfs in Γ contain individual constants, then there is a model for Γ whose domain is the set D of individual constants that occur in wfs of Γ.

Proof

Define a model M with domain D by specifying that the interpretation of any predicate letter A_k^n in Γ contains an n-tuple $\langle b_1, \ldots, b_n \rangle$ if and only if $A_k^n(b_1, \ldots, b_n)$ is in Γ. By induction on the number of connectives and quantifiers in any closed wf $\mathscr{E}$, it is easy to prove: (i) if $\mathscr{E}$ is in Γ, then $\mathscr{E}$ is true in M; and (ii) if $\neg\mathscr{E}$, is in Γ, then $\mathscr{E}$ is false in M. Hence, M is a model for Γ.

If a branch of a semantic tree remains open, the set Γ of wfs of that branch satisfies the hypotheses of Proposition 2.47. If follows that, if a branch of a semantic tree remains open, then the set Γ of wfs of that branch has a model M whose domain is the set of individual constants that appear in that branch. This yields half of the basic principle of semantic trees.

PROPOSITION 2.48

If all the branches of a semantic tree are closed, then the wf $\mathscr{B}$ at the root of the tree is unsatisfiable.

Proof

From the derivation rules it is clear that, if a sequence of wfs starts at $\mathscr{B}$ and continues down the tree through the applications of the rules, and if the wfs in that sequence are simultaneously satisfiable in some model M, then that sequence can be extended by another application of a rule so that the added wf(s) would also be true in M. Otherwise, the sequence would form an unclosed branch, contrary to our hypothesis. Assume now that $\mathscr{B}$ is satisfiable in a model M. Then, starting with $\mathscr{B}$, we could construct an infinite branch in which all the wfs are true in M. (In the case of a branching rule, if there are two ways to extend the sequence, we choose the left-hand wf.) Therefore, the branch would not be closed, contrary to our hypothesis. Hence, $\mathscr{B}$ is unsatisfiable.

This completes the proof of the basic principle of semantic trees. Notice that this principle does not yield a decision procedure for logical validity. If a closed wf $\mathscr{B}$ is not logically valid, the semantic tree of $\neg\mathscr{B}$ may (and often does) contain an infinite unclosed branch. At any stage of the construction of this tree, we have no general procedure for deciding whether or not, at some later stage, all branches of the tree will have become closed. Thus, we have no general way of knowing whether $\mathscr{B}$ is unsatisfiable.

For the sake of brevity, our exposition has been loose and imprecise. A clear and masterful study of semantic trees and related matters can be found in Smullyan (1968).

2.16 QUANTIFICATION THEORY ALLOWING EMPTY DOMAINS

Our definition in Section 2.2 of *interpretations* of a language assumed that the domain of an interpretation is non-empty. This was done for the sake of simplicity. If we allow the empty domain, questions arise as to the right way of defining the truth of a formula in such a domain.† Once that is decided, the corresponding class of valid formulas (that is, formulas true in all interpretations, including the one with an empty domain) becomes smaller, and it is difficult to find an axiom system that will have all such formulas as its theorems. Finally, an interpretation with an empty domain has little or no importance in applications of logic.

Nevertheless, the problem of finding a suitable treatment of such a more inclusive logic has aroused some curiosity and we shall present one possible approach. In order to do so, we shall have to restrict the scope of the investigation in the following ways.

First, our languages will contain no individual constants or function letters. The reason for this restriction is that it is not clear how to interpret individual constants or function letters when the domain of the interpretation is empty. Moreover, in first-order theories with equality, individual constants and function letters always can be replaced by new predicate letters, together with suitable axioms.‡

Second, we shall take every formula of the form $(\forall x)\mathscr{B}(x)$ to be true in the empty domain. This is based on parallelism with the case of a non-empty domain. To say that $(\forall x)\mathscr{B}(x)$ holds in a non-empty domain D amounts to asserting

$$(*) \quad \text{for any object } c, \text{if } c \in D, \text{then } \mathscr{B}(c)$$

When D is empty, '$c \in D$' is false and, therefore, 'if $c \in D$, then $\mathscr{B}(c)$' is true. Since this holds for arbitrary c, $(*)$ is true in the empty domain D, that is, $(\forall x)\mathscr{B}(x)$ is true in an empty domain. Not unexpectedly, $(\exists x)\mathscr{B}(x)$ will be false in an empty domain, since $(\exists x)\mathscr{B}(x)$ is equivalent to $\neg(\forall x)\neg\mathscr{B}(x)$.

These two conventions enable us to calculate the truth value of any closed formula in an empty domain. Every such formula is a truth-functional combination of formulas of the form $(\forall x)\mathscr{B}(x)$. Replace every subformula

†For example, should a formula of the form $(\forall x)(A_1^1(x) \wedge \neg A_1^1(x))$ be considered true in the empty domain?

‡For example, an individual constant b can be replaced by a new monadic predicate letter P, together with the axiom $(\exists y)(\forall x)(P(x) \Leftrightarrow x = y)$. Any axiom $\mathscr{B}(b)$ should be replaced by $(\forall x)(P(x) \Rightarrow \mathscr{B}(x))$.

$(\forall x)\mathscr{B}(x)$ by the truth value T and then compute the truth value of the whole formula.

It is not clear how we should define the truth value in the empty domain of a formula containing free variables. We might imitate what we do in the case of non-empty domains and take such a formula to have the same truth values as its universal closure. Since the universal closure is automatically true in the empty domain, this would have the uncomfortable consequence of declaring the formula $A_1^1(x) \wedge \neg A_1^1(x)$ to be true in the empty domain. For this reason, we shall confine our attention to sentences, that is, formulas without free variables.

A sentence will be said to be *inclusively valid* if it is true in all interpretations, including the interpretation with an empty domain. Every inclusively valid sentence is logically valid, but the converse does not hold. To see this, let $\mathbf{f}$ stand for a sentence $\mathscr{C} \wedge \neg\mathscr{C}$, where $\mathscr{C}$ is some fixed sentence. Now, $\mathbf{f}$ is false in the empty domain but $(\forall x)\mathbf{f}$ is true in the empty domain (since it begins with a universal quantifier). Thus the sentence $(\forall x)\mathbf{f} \Rightarrow \mathbf{f}$ is false in the empty domain and, therefore, not inclusively valid. However, it is logically valid, since every formula of the form $(\forall x)\mathscr{B} \Rightarrow \mathscr{B}$ is logically valid.

The problem of determining the inclusive validity of a sentence is reducible to that of determining its logical validity, since we know how to determine whether a sentence is true in the empty domain. Since the problem of determining logical validity will turn out to be unsolvable (by Proposition 3.54), the same applies to inclusive validity.

Now let us turn to the problem of finding an axiom system whose theorems are the inclusively valid sentences. We shall adapt for this purpose an axiom system $\mathrm{PP}^{\#}$ based on Exercise 2.28. As axioms we take all the following formulas (see the Logical Axioms on p. 69):

(A1) $\mathscr{B} \Rightarrow (\mathscr{C} \Rightarrow \mathscr{B})$

(A2) $(\mathscr{B} \Rightarrow (\mathscr{C} \Rightarrow \mathscr{D})) \Rightarrow ((\mathscr{B} \Rightarrow \mathscr{C}) \Rightarrow (\mathscr{B} \Rightarrow \mathscr{D}))$

(A3) $(\neg\mathscr{C} \Rightarrow \neg\mathscr{B}) \Rightarrow ((\neg\mathscr{C} \Rightarrow \mathscr{B}) \Rightarrow \mathscr{C})$

(A4) $(\forall x)\mathscr{B}(x) \Rightarrow \mathscr{B}(y)$ if $\mathscr{B}(x)$ is a wf of $\mathscr{L}$ and y is a variable that is free for x in $\mathscr{B}(x)$. (Recall that, if y is x itself, then the axiom has the form $(\forall x)\mathscr{B} \Rightarrow \mathscr{B}$. In addition, x need not be free in $\mathscr{B}(x)$.)

(A5) $(\forall x)(\mathscr{B} \Rightarrow \mathscr{C}) \Rightarrow (\mathscr{B} \Rightarrow (\forall x)\mathscr{C})$ if $\mathscr{B}$ contains no free occurrences of x.

(A6) $(\forall y_1)\ldots(\forall y_n)(\mathscr{B} \Rightarrow \mathscr{C}) \Rightarrow [(\forall y_1)\ldots(\forall y_n)\mathscr{B} \Rightarrow (\forall y_1)\ldots(\forall y_n)\mathscr{C}]$

together with all formulas obtained by prefixing any sequence of universal quantifiers to instances of (A1)–(A6).

Modus ponens (MP) will be the only rule of inference.

PP denotes the pure first-order predicate calculus, whose axioms are (A1)–(A5), whose rules of inference are MP and Gen, and whose language contains no individual constants or function letters. By Gödel's completeness theorem (Corollary 2.19), the theorems of PP are the same as the

logically valid formulas in PP. Exercise 2.28 shows first that Gen is a derived rule of inference of $PP^{\#}$, that is, if $\vdash_{PP^{\#}} \mathscr{D}$, then $\vdash_{PP^{\#}} (\forall x)\mathscr{D}$, and second that PP and $PP^{\#}$ have the same theorems. Hence, the theorems of $PP^{\#}$ are the logically valid formulas.

Let $PPS^{\#}$ be the same system as $PP^{\#}$ except that, as axioms, we take only the axioms of $PP^{\#}$ that are sentences. Since MP takes sentences into sentences, all theorems of $PPS^{\#}$ are sentences. Since all axioms of $PPS^{\#}$ are axioms of $PP^{\#}$, all theorems of $PPS^{\#}$ are logically valid sentences. Let us show that the converse holds.

PROPOSITION 2.49

Every logically valid sentence is a theorem of $PPS^{\#}$.

Proof

Let $\mathscr{B}$ be any logically valid sentence. We know that $\mathscr{B}$ is a theorem of $PP^{\#}$. Let us show that $\mathscr{B}$ is a theorem of $PPS^{\#}$. In a proof of $\mathscr{B}$ in $PP^{\#}$, let $u_1, \ldots, u_n$ be the free variables (if any) in the proof, and prefix $(\forall u_1) \ldots (\forall u_n)$ to all steps of the proof. Then each step goes into a theorem of $PPS^{\#}$. To see this, first note that axioms of $PP^{\#}$ go into axioms of $PPS^{\#}$. Second, assume that $\mathscr{D}$ comes from $\mathscr{C}$ and $\mathscr{C} \Rightarrow \mathscr{D}$ by MP in the original proof and that $(\forall u_1) \ldots (\forall u_n)\mathscr{C}$ and $(\forall u_1) \ldots (\forall u_n)(\mathscr{C} \Rightarrow \mathscr{D})$ are provable in $PPS^{\#}$. Since $(\forall u_1) \ldots (\forall u_n)(\mathscr{C} \Rightarrow \mathscr{D}) \Rightarrow [(\forall u_1) \ldots (\forall u_n)\mathscr{C} \Rightarrow (\forall u_1) \ldots (\forall u_n)\mathscr{D}]$ is an instance of axiom (A6) of $PPS^{\#}$, it follows that $(\forall u_1) \ldots (\forall u_n)\mathscr{D}$ is provable in $PPS^{\#}$. Thus, $(\forall u_1) \ldots (\forall u_n)\mathscr{B}$ is a theorem of $PPS^{\#}$. Then n applications of axiom (A4) and MP show that $\mathscr{B}$ is a theorem of $PPS^{\#}$.

Not all axioms of $PPS^{\#}$ are inclusively valid. For example, the sentence $(\forall x)\mathbf{f} \Rightarrow \mathbf{f}$ discussed earlier is an instance of axiom (A4) that is not inclusively valid. So, in order to find an axiom system for inclusive validity, we must modify $PPS^{\#}$.

If P is a sequence of variables $u_1, \ldots, u_n$, then by $\forall P$ we shall mean the expression $(\forall u_1) \ldots (\forall u_n)$.

Let the axiom system ETH be obtained from $PPS^{\#}$ by changing axiom (A4) into:

(A4′) All sentences of the form $\forall P[(\forall x)\mathscr{B}(x) \Rightarrow \mathscr{B}(y)]$, where y is free for x in $\mathscr{B}(x)$ and x is free in $\mathscr{B}(x)$, and P is a sequence of variables that includes all variables free in $\mathscr{B}$ (and possibly others).

MP is the only rule of inference.

It is obvious that all axioms of ETH are inclusively valid.

LEMMA 2.50

If $\mathscr{T}$ is an instance of a tautology and P is a sequence of variables that contains all free variables in $\mathscr{T}$, then $\vdash_{\text{ETH}} \forall P\mathscr{T}$.

Proof

By the completeness of axioms (A1)–(A3) for the propositional calculus, there is a proof of $\mathscr{T}$ using MP and instances of (A1)–(A3). If we prefix $\forall P$ to all steps of that proof, the resulting sentences are all theorems of ETH. In the case when an original step $\mathscr{B}$ was an instance of (A1)-(A3), $\forall P\mathscr{B}$ is an axiom of ETH. For steps that result from MP, we use axiom (A6).

LEMMA 2.51

If P is a sequence of variables that includes all free variables of $\mathscr{B} \Rightarrow \mathscr{C}$, and $\vdash_{\text{ETH}} \forall P\mathscr{B}$ and $\vdash_{\text{ETH}} \forall P[\mathscr{B} \Rightarrow \mathscr{C}]$, then $\vdash_{\text{ETH}} \forall P\mathscr{C}$.

Proof

Use axiom (A6) and MP.

LEMMA 2.52

If P is a sequence of variables that includes all free variables of $\mathscr{B}, \mathscr{C}, \mathscr{D}$, and $\vdash_{\text{ETH}} \forall P[\mathscr{B} \Rightarrow \mathscr{C}]$ and $\vdash_{\text{ETH}} \forall P[\mathscr{C} \Rightarrow \mathscr{D}]$, then $\vdash_{\text{ETH}} \forall P[\mathscr{B} \Rightarrow \mathscr{D}]$.

Proof

Use the tautology $(\mathscr{B} \Rightarrow \mathscr{C}) \Rightarrow ((\mathscr{C} \Rightarrow \mathscr{D}) \Rightarrow (\mathscr{B} \Rightarrow \mathscr{D}))$, Lemma 2.50, and Lemma 2.51 twice.

LEMMA 2.53

If x is not free in $\mathscr{B}$ and P is a sequence of variables that contains all free variables of $\mathscr{B}$, $\vdash_{\text{ETH}} \forall P[\mathscr{B} \Rightarrow (\forall x)\mathscr{B}]$.

Proof

By axiom (A5), $\vdash_{\text{ETH}} \forall P[(\forall x)(\text{B} \Rightarrow \text{B}) \Rightarrow (\text{B} \Rightarrow (\forall x)\text{B})]$. By Lemma 2.50, $\vdash_{\text{ETH}} \forall P[(\forall x)(\mathscr{B} \Rightarrow \mathscr{B})]$. Now use Lemma 2.51.

COROLLARY 2.54

If $\mathscr{B}$ has no free variables, then $\vdash_{\text{ETH}} \mathscr{B} \Rightarrow (\forall x)\mathscr{B}$.

LEMMA 2.55

If x is not free in $\mathscr{B}$ and P is a sequence of variables that includes all variables free in $\mathscr{B}$, then $\vdash_{\text{ETH}} \forall P[\neg(\forall x)\mathbf{f} \Rightarrow ((\forall x)\mathscr{B} \Rightarrow \mathscr{B})]$.

Proof

$\vdash_{\text{ETH}} \forall P[\neg\mathscr{B} \Rightarrow (\mathscr{B} \Rightarrow \mathbf{f})]$ by Lemma 2.50. By Lemma 2.53, $\vdash_{\text{ETH}} \forall P$ $[(\mathscr{B} \Rightarrow \mathbf{f}) \Rightarrow (\forall x)(\mathscr{B} \Rightarrow \mathbf{f})]$. Hence, by Lemma 2.52, $\vdash_{\text{ETH}} \forall P[\neg\mathscr{B} \Rightarrow (\forall x)$ $(\mathscr{B} \Rightarrow \mathbf{f})]$. By axiom (A6), $\vdash_{\text{ETH}} \forall P[(\forall x)(\mathscr{B} \Rightarrow \mathbf{f}) \Rightarrow ((\forall x)\mathscr{B} \Rightarrow (\forall x)\mathbf{f})]$. Hence, by Lemma 2.52, $\vdash_{\text{ETH}} \forall P[\neg\mathscr{B} \Rightarrow ((\forall x)\mathscr{B} \Rightarrow (\forall x)\mathbf{f})]$. Since $[\neg\mathscr{B} \Rightarrow$ $((\forall x)\mathscr{B} \Rightarrow (\forall x)\mathbf{f})] \Rightarrow [\neg(\forall x)\mathbf{f} \Rightarrow ((\forall x)\mathscr{B} \Rightarrow \mathscr{B})]$ is an instance of a tautology, Lemmas 2.50 and 2.51 yield $\vdash_{\text{ETH}} \forall P[\neg(\forall x)\mathbf{f} \Rightarrow ((\forall x)\mathscr{B} \Rightarrow \mathscr{B})]$.

PROPOSITION 2.56

ETH + $\{\neg(\forall x)\mathbf{f}\}$ is a complete axiom system for logical validity, that is, a sentence is logically valid if and only if it is a theorem of the system.

Proof

All axioms of the system are logically valid. (Note that $(\forall x)\mathbf{f}$ is false in all interpretations with a non-empty domain and, therefore, $\neg(\forall x)\mathbf{f}$ is true in all such domains.) By Proposition 2.49, all logically valid sentences are provable in $\text{PPS}^{\#}$. The only axioms of $\text{PPS}^{\#}$ missing from ETH are those of the form $\forall P[(\forall x)\mathscr{B} \Rightarrow \mathscr{B}]$, where x is not free in $\mathscr{B}$ and P is any sequence of variables that include all free variables of $\mathscr{B}$. By Lemma 2.55, $\vdash_{\text{ETH}} \forall P[\neg(\forall x)\mathbf{f} \Rightarrow ((\forall x)\mathscr{B} \Rightarrow \mathscr{B})]$. By Corollary 2.54, $\forall P[\neg(\forall x)\mathbf{f}]$ will be derivable in ETH + $\{\neg(\forall x)\mathbf{f}\}$. Hence, $\forall P[(\forall x)\mathscr{B} \Rightarrow \mathscr{B}]$ is obtained by using axiom (A6).

LEMMA 2.57

If P is a sequence of variables that include all free variables of $\mathscr{B}$, $\vdash_{\text{ETH}} \forall P[(\forall x)\mathbf{f} \Rightarrow ((\forall x)\mathscr{B} \Leftrightarrow \mathbf{t})]$, where $\mathbf{t}$ is $\neg\mathbf{f}$.

Proof

Since $\mathbf{f} \Rightarrow \mathscr{B}$ is an instance of a tautology, Lemma 2.50 yields $\vdash_{\text{ETH}} \forall P(\forall x)[\mathbf{f} \Rightarrow \mathscr{B}]$. By axiom (A6), $\vdash_{\text{ETH}} \forall P[(\forall x)[\mathbf{f} \Rightarrow \mathscr{B}] \Rightarrow [(\forall x)\mathbf{f} \Rightarrow (\forall x)\mathscr{B}]]$. Hence, $\vdash_{\text{ETH}} \forall P[(\forall x)\mathbf{f} \Rightarrow (\forall x)\mathscr{B}]$ by Lemma 2.51. Since $(\forall x)\mathscr{B} \Rightarrow [(\forall x)\mathscr{B} \Leftrightarrow \mathbf{t}]$ is an instance of a tautology, Lemma 2.50 yields $\vdash_{\text{ETH}} \forall P[(\forall x)\mathscr{B} \Rightarrow [(\forall x)\mathscr{B} \Leftrightarrow \mathbf{t}]]$. Now, by Lemma 2.52, $\vdash_{\text{ETH}} \forall P[(\forall x)\mathbf{f} \Rightarrow [(\forall x)\mathscr{B} \Leftrightarrow \mathbf{t}]]$.

Given a formula $\mathscr{B}$, construct a formula $\mathscr{B}^*$ in the following way. Moving from left to right, replace each universal quantifier and its scope by $\mathbf{t}$.

LEMMA 2.58

If P is a sequence of variables that include all free variables of $\mathscr{B}$, then $\vdash_{\text{ETH}} \forall P[(\forall x)\mathbf{f} \Rightarrow [\mathscr{B} \Leftrightarrow \mathscr{B}^*]]$.

Proof

Apply Lemma 2.57 successively to the formulas obtained in the stepwise construction of $\mathscr{B}^*$. We leave the details to the reader.

PROPOSITION 2.59

ETH is a complete axiom system for inclusive validity, that is, a sentence $\mathscr{B}$ is inclusively valid if and only if it is a theorem of ETH.

Proof

Assume $\mathscr{B}$ is a sentence valid for all interpretations. We must show that $\vdash_{\text{ETH}} \mathscr{B}$. Since $\mathscr{B}$ is valid in all non-empty domains, Proposition 2.56 implies that $\mathscr{B}$ is provable in ETH + $\{\neg(\forall x)\mathbf{f}\}$. Hence, by the deduction theorem,

$$(+) \quad \vdash_{\text{ETH}} \neg(\forall x)\mathbf{f} \Rightarrow \mathscr{B}.$$

Now, by Lemma 2.58,

$$(\%) \quad \vdash_{\text{ETH}} (\forall x)\mathbf{f} \Rightarrow [\mathscr{B} \Leftrightarrow \mathscr{B}^*]$$

(Since $\mathscr{B}$ has no free variables, we can take P in Lemma 2.58 to be empty.) Hence, $[(\forall x)\mathbf{f} \Rightarrow [\mathscr{B} \Leftrightarrow \mathscr{B}^*]]$ is valid for all interpretations. Since $(\forall x)\mathbf{f}$ is valid in the empty domain and $\mathscr{B}$ is valid for all interpretations, $\mathscr{B}^*$ is valid in the empty domain. But $\mathscr{B}^*$ is a truth-functional combination of $\mathbf{t}$s.

So, $\mathscr{B}^*$ must be truth-functionally equivalent to either **t** or **f**. Since it is valid in the empty domain, it is truth-functionally equivalent to **t**. Hence, $\vdash_{\text{ETH}} \mathscr{B}^*$. Therefore by (%), $\vdash_{\text{ETH}} (\forall x)\mathbf{f} \Rightarrow \mathscr{B}$. This, together with (+), yields $\vdash_{\text{ETH}} \mathscr{B}$.

The ideas and methods used in this section stem largely, but not entirely, from a paper by Hailperin (1953).† That paper also made use of an idea in Mostowski (1951b), the idea that underlies the proof of Proposition 2.59. Mostowski's approach to the logic of the empty domain is quite different from Hailperin's and results in a substantially different axiom system for inclusive validity. For example, when $\mathscr{B}$ does not contain x free, Mostowski interprets $(\forall x)\mathscr{B}$ and $(\exists x)\mathscr{B}$ to be $\mathscr{B}$ itself. This makes $(\forall x)\mathbf{f}$ equivalent to **f**, rather than to **t**, as in our development.

†The name ETH comes from 'empty domain' and 'Theodore Hailperin'. My simplification of Hailperin's axiom system was suggested by a similar simplification in Quine (1954).

3 Formal Number Theory

3.1 AN AXIOM SYSTEM

Together with geometry, the theory of numbers is the most immediately intuitive of all branches of mathematics. It is not surprising, then, that attempts to formalize mathematics and to establish a rigorous foundation for mathematics should begin with number theory. The first semi-axiomatic presentation of this subject was given by Dedekind in 1879 and, in a slightly modified form, has come to be known as Peano's postulates.[†] It can be formulated as follows:

(P1) 0 is a natural number.[‡]
(P2) If x is a natural number, there is another natural number denoted by x' (and called the *successor* of x).[§]
(P3) $0 \neq x'$ for every natural number x.
(P4) If $x' = y'$, then $x = y$.
(P5) If Q is a property that may or may not hold for any given natural number, and if (I) 0 has the property Q and (II) whenever a natural number x has the property Q, then x' has the property Q, then all natural numbers have the property Q (mathematical induction principle).

These axioms, together with a certain amount of set theory, can be used to develop not only number theory but also the theory of rational, real and complex numbers (see Mendelson, 1973). However, the axioms involve certain intuitive notions, such as 'property', that prevent this system from being a rigorous formalization. We therefore shall build a first-order theory S that is based upon Peano's postulates and seems to be adequate for the proofs of all the basic results of elementary number theory.

The language $\mathscr{L}_A$ of our theory S will be called the *language of arithmetic*. $\mathscr{L}_A$ has a single predicate letter A_1^2. As usual, we shall write $t = s$ for $A_1^2(t, s)$. $\mathscr{L}_A$ has one individual constant a_1. We shall use 0 as an alternative notation

[†]For historical information, see Wang (1957).
[‡]The natural numbers are supposed to be the non-negative integers $0, 1, 2, \ldots$.
[§]The intuitive meaning of x' is $x + 1$.

for a_1. Finally, $\mathscr{L}_A$ has three function letters, f_1^1, f_1^2 and f_2^2. We shall write (t') instead of $f_1^1(t)$, $(t+s)$ instead of $f_1^2(t,s)$, and $(t \cdot s)$ instead of $f_2^2(t,s)$. However, we shall write $t', t+s$, and $t \cdot s$ instead of $(t'), (t+s)$, and $(t \cdot s)$ whenever this will cause no confusion.

The proper axioms of S are:

(S1) $x_1 = x_2 \Rightarrow (x_1 = x_3 \Rightarrow x_2 = x_3)$
(S2) $x_1 = x_2 \Rightarrow x_1' = x_2'$
(S3) $0 \neq x_1'$
(S4) $x_1' = x_2' \Rightarrow x_1 = x_2$
(S5) $x_1 + 0 = x_1$
(S6) $x_1 + x_2' = (x_1 + x_2)'$
(S7) $x_1 \cdot 0 = 0$
(S8) $x_1 \cdot (x_2)' = (x_1 \cdot x_2) + x_1$
(S9) $\mathscr{B}(0) \Rightarrow ((\forall x)(\mathscr{B}(x) \Rightarrow \mathscr{B}(x')) \Rightarrow (\forall x)\mathscr{B}(x))$ for any wf $\mathscr{B}(x)$ of S.

We shall call (S9) the *principle of mathematical induction*. Notice that axioms (S1)–(S8) are particular wfs, whereas (S9) is an axiom schema providing an infinite number of axioms.†

Axioms (S3) and (S4) correspond to Peano postulates (P3) and (P4), respectively. Peano's axioms (P1) and (P2) are taken care of by the presence of 0 as an individual constant and f_1^1 as a function letter. Our axioms (S1) and (S2) furnish some needed properties of equality; they would have been assumed as intuitively obvious by Dedekind and Peano. Axioms (S5)–(S8) are the recursion equations for addition and multiplication. They were not assumed by Dedekind and Peano because the existence of operations $+$ and $\cdot$ satisfying (S5)–(S8) is derivable by means of intuitive set theory, which was presupposed as a background theory (see Mendelson, 1973, chapter 2, Theorems 3.1 and 5.1).

Any theory that has the same theorems as S is often referred to in the literature as *Peano arithmetic,* or simply PA.

From (S9) by MP, we can obtain the *induction rule*:

$$\mathscr{B}(0), (\forall x)(\mathscr{B}(x) \Rightarrow \mathscr{B}(x')) \vdash_S (\forall x)\mathscr{B}(x).$$

It will be our immediate aim to establish the usual rules of equality; that is, we shall show that the properties (A6) and (A7) of equality (see page 95) are derivable in S and, hence, that S is a first-order theory with equality.

First, for convenience and brevity in carrying out proofs, we cite some immediate, trivial consequences of the axioms.

†However, (S9) cannot fully correspond to Peano's postulate (P5), since the latter refers intuitively to the $2^{\aleph_0}$ properties of natural numbers, whereas (S9) can take care of only the denumberable number of properties defined by wfs of $\mathscr{L}_A$.

LEMMA 3.1

For any terms t, s, r of $\mathscr{L}_A$, the following wfs are theorems of S.

(S1′) $t = r \Rightarrow (t = s \Rightarrow r = s)$
(S2′) $t = r \Rightarrow t' = r'$
(S3′) $0 \neq t'$
(S4′) $t' = r' \Rightarrow t = r$
(S5′) $t + 0 = t$
(S6′) $t + r' = (t + r)'$
(S7′) $t \cdot 0 = 0$
(S8′) $t \cdot r' = (t \cdot r) + t$

Proof

(S1′)–(S8′) follow from (S1)–(S8), respectively. First form the closure by means of Gen, use Exercise 2.48 to change all the bound variables to variables not occuring in terms t, r, s, and then apply rule A4 with the appropriate terms t, r, s.†

PROPOSITION 3.2.

For any terms t, s, r, the following wfs are theorems of S.

(a) $t = t$
(b) $t = r \Rightarrow r = t$
(c) $t = r \Rightarrow (r = s \Rightarrow t = s)$
(d) $r = t \Rightarrow (s = t \Rightarrow r = s)$
(e) $t = r \Rightarrow t + s = r + s$
(f) $t = 0 + t$
(g) $t' + r = (t + r)'$
(h) $t + r = r + t$
(i) $t = r \Rightarrow s + t = s + r$
(j) $(t + r) + s = t + (r + s)$
(k) $t = r \Rightarrow t \cdot s = r \cdot s$
(l) $0 \cdot t = 0$
(m) $t' \cdot r = t \cdot r + r$

†The change of bound variables is necessary in some cases. For example, if we want to obtain $x_2 = x_1 \Rightarrow x_2' = x_1'$ from $x_1 = x_2 \Rightarrow x_1' = x_2'$, we first obtain $(\forall x_1)(\forall x_2)(x_1 = x_2 \Rightarrow x_1' = x_2')$. We cannot apply rule A4 to drop $(\forall x_1)$ and replace x_1 by x_2, since x_2 is not free for x_1 in $(\forall x_2)(x_1 = x_2 \Rightarrow x_1' = x_2')$. From now on, we shall assume without explicit mention that the reader is aware that we sometimes have to change bound variables when we use Gen and rule A4.

(n) $t \cdot r = r \cdot t$
(o) $t = r \Rightarrow s \cdot t = s \cdot r$

Proof

(a) 1.	$t + 0 = t$	(S5′)
2.	$(t + 0 = t) \Rightarrow (t + 0 = t \Rightarrow t = t)$	(S1′)
3.	$t + 0 = t \Rightarrow t = t$	1, 2, MP
4.	$t = t$	1, 3, MP
(b) 1.	$t = r \Rightarrow (t = t \Rightarrow r = t)$	(S1′)
2.	$t = t \Rightarrow (t = r \Rightarrow r = t)$	1, tautology, MP
3.	$t = r \Rightarrow r = t$	2, part (a), MP
(c) 1.	$r = t \Rightarrow (r = s \Rightarrow t = s)$	(S1′)
2.	$t = r \Rightarrow r = t$	Part (b)
3.	$t = r \Rightarrow (r = s \Rightarrow t = s)$	1, 2, tautology, MP
(d) 1.	$r = t \Rightarrow (t = s \Rightarrow r = s)$	Part (c)
2.	$t = s \Rightarrow (r = t \Rightarrow r = s)$	1, tautology, MP
3.	$s = t \Rightarrow t = s$	Part (b)
4.	$s = t \Rightarrow (r = t \Rightarrow r = s)$	2, 3, tautology, MP

(e) Apply the induction rule to $\mathscr{B}(z) : x = y \Rightarrow x + z = y + z$.

(i) 1.	$x + 0 = x$	(S5′)
2.	$y + 0 = y$	(S5′)
3.	$x = y$	Hyp
4.	$x + 0 = y$	1, 3, part (c), MP
5.	$x + 0 = y + 0$	4, 2, part (d), MP
6.	$\vdash_S x = y \Rightarrow x + 0 = y + 0$	1–5, deduction theorem

Thus, $\vdash_S \mathscr{B}(0)$.

(ii) 1.	$x = y \Rightarrow x + z = y + z$	Hyp
2.	$x = y$	Hyp
3.	$x + z' = (x + z)'$	(S6′)
4.	$y + z' = (y + z)'$	(S6′)
5.	$x + z = y + z$	1, 2, MP
6.	$(x + z)' = (y + z)'$	5, (S2′), MP
7.	$x + z' = (y + z)'$	3, 6, part (c), MP
8.	$x + z' = y + z'$	4, 7, part (d), MP
9.	$\vdash_S (x = y \Rightarrow x + z = y + z) \Rightarrow (x = y \Rightarrow x + z' = y + z')$	1–8, deduction theorem twice

Thus, $\vdash_S \mathscr{B}(z) \Rightarrow \mathscr{B}(z')$, and, by Gen, $\vdash_S (\forall z)(\mathscr{B}(z) \Rightarrow \mathscr{B}(z'))$. Hence, $\vdash_S (\forall z)\mathscr{B}(z)$ by the induction rule. Therefore, by Gen and rule A4, $\vdash_S t = r \Rightarrow t + s = r + s$.

(f) Let $\mathscr{B}(x)$ be $x = 0 + x$.

(i) $\vdash_S 0 = 0 + 0$ by (S5′), part (b) and MP; thus, $\vdash_S \mathscr{B}(0)$.

(ii)

1.	$x = 0 + x$	Hyp
2.	$0 + x' = (0 + x)'$	(S6′)
3.	$x' = (0 + x)'$	1, (S2′), MP
4.	$x' = 0 + x'$	3, 2, part (d), MP
5.	$\vdash_S x = 0 + x \Rightarrow x' = 0 + x'$	1–4, deduction theorem

Thus, $\vdash_S \mathscr{B}(x) \Rightarrow \mathscr{B}(x')$ and, by Gen, $\vdash_S (\forall x)(\mathscr{B}(x) \Rightarrow \mathscr{B}(x'))$. So, by (i), (ii) and the induction rule, $\vdash_S (\forall x)(x = 0 + x)$, and then, by rule A4, $\vdash_S t = 0 + t$.

(g) Let $\mathscr{B}(y)$ be $x' + y = (x + y)'$.

(i)

1.	$x' + 0 = x'$	(S5′)
2.	$x + 0 = x$	(S5′)
3.	$(x + 0)' = x'$	2, (S2′), MP
4.	$x' + 0 = (x + 0)'$	1, 3, part (d), MP

Thus, $\vdash_S \mathscr{B}(0)$.

(ii)

1.	$x' + y = (x + y)'$	Hyp
2.	$x' + y' = (x' + y)'$	(S6′)
3.	$(x' + y)' = (x + y)''$	1, (S2′), MP
4.	$x' + y' = (x + y)''$	2, 3, part (c), MP
5.	$x + y' = (x + y)'$	(S6′)
6.	$(x + y')' = (x + y)''$	5, (S2′), MP
7.	$x' + y' = (x + y')'$	4, 6, part (d), MP
8.	$\vdash_S x' + y = (x + y)' \Rightarrow x' + y' = (x + y')'$	1–7, deduction theorem

Thus, $\vdash_S \mathscr{B}(y) \Rightarrow \mathscr{B}(y')$, and, by Gen, $\vdash_S (\forall y)(\mathscr{B}(y) \Rightarrow \mathscr{B}(y'))$. Hence, by (i), (ii) and the induction rule, $\vdash_S (\forall y)(x' + y = (x + y)')$. By Gen and rule A4, $\vdash_S t' + r = (t + r)'$.

(h) Let $\mathscr{B}(y)$ be $x + y = y + x$.

(i)

1.	$x + 0 = x$	(S5′)
2.	$x = 0 + x$	Part (f)
3.	$x + 0 = 0 + x$	1, 2, part (c), MP

Thus, $\vdash_S \mathscr{B}(0)$.

(ii)

1.	$x + y = y + x$	Hyp
2.	$x + y' = (x + y)'$	(S6′)
3.	$y' + x = (y + x)'$	Part (g)
4.	$(x + y)' = (y + x)'$	1, (S2′), MP
5.	$x + y' = (y + x)'$	2, 4, part (c), MP
6.	$x + y' = y' + x$	5, 3, part (d), MP
7.	$\vdash_S x + y = y + x \Rightarrow x + y' = y' + x$	1–6, deduction theorem

Thus, $\vdash_S \mathscr{B}(y) \Rightarrow \mathscr{B}(y')$ and, by Gen, $\vdash_S (\forall y)(\mathscr{B}(y) \Rightarrow \mathscr{B}(y'))$. So, by (i), (ii) and the induction rule, $\vdash_S (\forall y)(x+y=y+x)$. Then, by rule A4, Gen and rule A4, $\vdash_S t+r=r+t$.

(i) 1. $t=r \Rightarrow t+s=r+s$	Part (e)
2. $t+s=s+t$	Part (h)
3. $r+s=s+r$	Part (h)
4. $t=r$	Hyp
5. $t+s=r+s$	1, 4, MP
6. $s+t=r+s$	2, 5, (S1′) MP
7. $s+t=s+r$	6, 3, part (c), MP
8. $\vdash_S t=r \Rightarrow s+t=s+r$	1–7, deduction theorem

(j) Let $\mathscr{B}(z)$ be $(x+y)+z=x+(y+z)$.

(i) 1. $(x+y)+0=x+y$	(S5′)
2. $y+0=y$	(S5′)
3. $x+(y+0)=x+y$	2, part (j), MP
4. $(x+y)+0=x+(y+0)$	1, 3, part (d), MP

Thus, $\vdash_S \mathscr{B}(0)$.

(ii) 1. $(x+y)+z=x+(y+z)$	Hyp
2. $(x+y)+z'=((x+y)+z)'$	(S6′)
3. $((x+y)+z)'=(x+(y+z))'$	1, (S2′), MP
4. $(x+y)+z'=(x+(y+z))'$	2, 3, part (c), MP
5. $y+z'=(y+z)'$	(S6′)
6. $x+(y+z')=x+(y+z)'$	5, part (i), MP
7. $x+(y+z)'=(x+(y+z))'$	(S6′)
8. $x+(y+z')=(x+(y+z))'$	6, 7, part (c), MP
9. $(x+y)+z'=x+(y+z')$	4, 8, part (d), MP
10. $\vdash_S (x+y)+z=x+(y+z) \Rightarrow (x+y)+z'=x+(y+z')$	1–9, deduction theorem

Thus, $\vdash_S \mathscr{B}(z) \Rightarrow \mathscr{B}(z')$ and, by Gen, $\vdash_S (\forall z)(\mathscr{B}(z) \Rightarrow (\mathscr{B}(z'))$. So, by (i), (ii) and the induction rule, $\vdash_S (\forall z)\mathscr{B}(z)$, and then, by Gen and rule A4, $\vdash_S (t+r)+s=t+(r+s)$.

Parts (k)–(o) are left as exercises.

COROLLARY 3.3

S is a theory with equality.

Proof

By Proposition 2.25, this reduces to parts (a) (e), (i), (k) and (o) of proposition 3.2, and (S2′).

Notice that the interpretation in which:

(a) the set of non-negative integers is the domain
(b) the integer 0 is the interpretation of the symbol 0
(c) the successor operation (addition of 1) is the interpretation of the $'$ function (that is, of f_1^1)
(d) ordinary addition and multiplication are the interpretations of $+$ and $\cdot$
(e) the interpretation of the predicate letter $=$ is the identity relation

is a normal model for S. This model is called the *standard interpretation* or *standard model.* Any normal model for S that is not isomorphic to the standard model will be called a *non-standard model* for S.

If we recognize the standard interpretation to be a model for S, then, of course, S is consistent. However, this kind of semantic argument, involving as it does a certain amount of set-theoretic reasoning, is regarded by some as too precarious to serve as a basis for consistency proofs. Moreover, we have not proved in a rigorous way that the axioms of S are true under the standard interpretation, but we have taken it as intuitively obvious. For these and other reasons, when the consistency of S enters into the argument of a proof, it is common practice to take the statement of the consistency of S as an explicit unproved assumption.

Some important additional properties of addition and multiplication are covered by the following result.

PROPOSITION 3.4

For any terms t, r, s, the following wfs are theorems of S.

(a) $t \cdot (r + s) = (t \cdot r) + (t \cdot s)$ (distributivity)
(b) $(r + s) \cdot t = (r \cdot t) + (s \cdot t)$ (distributivity)
(c) $(t \cdot r) \cdot s = t \cdot (r \cdot s)$ (associativity of $\cdot$)
(d) $t + s = r + s \Rightarrow t = r$ (cancellation law for $+$)

Proof

(a) Prove $\vdash_S x \cdot (y + z) = (x \cdot y) + (x \cdot z)$ by induction on z.
(b) Use part (a) and Proposition 3.2(n).
(c) Prove $\vdash_S (x \cdot y) \cdot z = x \cdot (y \cdot z)$ by induction on z.
(d) Prove $\vdash_S x + z = y + z \Rightarrow x = y$ by induction on z. This requires, for the first time, use of (S4′).

The terms $0, 0', 0'', 0''', \ldots$ we shall call *numerals* and denote by $\overline{0}, \overline{1}, \overline{2}, \overline{3}, \ldots$. More precisely, $\overline{0}$ is 0 and, for any natural number n, $\overline{n+1}$ is $(\overline{n})'$. In general, if n is a natural number, $\overline{n}$ stands for the numeral consisting of 0 followed by n strokes. The numerals can be defined recursively by stating that 0 is a numeral and, if u is a numeral, then u' is also a numeral.

PROPOSITION 3.5

The following are theorems of S.

(a) $t+\bar{1}=t'$
(b) $t\cdot\bar{1}=t$
(c) $t\cdot\bar{2}=t+t$
(d) $t+s=0\Rightarrow t=0\wedge s=0$
(e) $t\neq 0\Rightarrow(s\cdot t=0\Rightarrow s=0)$
(f) $t+s=\bar{1}\Rightarrow(t=0\wedge s=\bar{1})\vee(t=\bar{1}\wedge s=0)$
(g) $t\cdot s=\bar{1}\Rightarrow(t=\bar{1}\wedge s=\bar{1})$
(h) $t\neq 0\Rightarrow(\exists y)(t=y')$
(i) $s\neq 0\Rightarrow(t\cdot s=r\cdot s\Rightarrow t=r)$
(j) $t\neq 0\Rightarrow(t\neq 1\Rightarrow(\exists y)(t=y''))$

Proof

(a)
1. $t+0'=(t+0)'$ (S6′)
2. $t+0=t$ (S5′)
3. $(t+0)'=t'$ 2, (S2′), MP
4. $t+0'=t'$ 1, 3, Proposition 3.2(c), MP
5. $t+\bar{1}=t'$ 4, abbreviation

(b)
1. $t\cdot 0'=t\cdot 0+t$ (S8′)
2. $t\cdot 0=0$ (S7′)
3. $t\cdot 0+t=0+t$ 2, Proposition 3.2(e), MP
4. $t\cdot 0'=0+t$ 1, 3, Proposition 3.2(c), MP
5. $0+t=t$ Proposition 3.2(f,b), MP
6. $t\cdot 0'=t$ 4, 5, Proposition 3.2(c), MP
7. $t\cdot\bar{1}=t$ 6, abbreviation

(c)
1. $t\cdot(\bar{1})'=(t\cdot\bar{1})+t$ (S8′)
2. $t\cdot\bar{1}=t$ Part (b)
3. $(t\cdot\bar{1})+t=t+t$ 2, Proposition 3.2(e), MP
4. $t\cdot(\bar{1})'=t+t$ 1, 3, Proposition 3.2(c), MP
5. $t\cdot\bar{2}=t+t$ 4, abbreviation

(d) Let $\mathscr{B}(y)$ be $x+y=0\Rightarrow x=0\wedge y=0$. It is easy to prove that $\vdash_S\mathscr{B}(0)$. Also, since $\vdash_S(x+y)'\neq 0$ by (S3′) and Proposition 3.2(b), it follows by (S6′) that $\vdash_S x+y'\neq 0$. Hence, $\vdash_S\mathscr{B}(y')$ by the tautology $\neg A\Rightarrow(A\Rightarrow B)$. So, $\vdash_S\mathscr{B}(y)\Rightarrow\mathscr{B}(y')$ by the tautology $A\Rightarrow(B\Rightarrow A)$. Then, by the induction rule, $\vdash_S(\forall y)\mathscr{B}(y)$ and then, by rule A4, Gen and rule A4, we obtain the theorem.

(e) The proof is similar to that for part (d) and is left as an exercise.

(f) Use induction on y in the wf $x+y=\bar{1}\Rightarrow((x=0\wedge y=\bar{1})\vee(x=\bar{1}\wedge y=0))$.

(g) Use induction on y in $x\cdot y=\bar{1}\Rightarrow(x=\bar{1}\wedge y=\bar{1})$.

(h) Perform induction on x in $x \neq 0 \Rightarrow (\exists w)(x = w')$.

(i) Let $\mathscr{B}(y)$ be $(\forall x)(z \neq 0 \Rightarrow (x \cdot z = y \cdot z \Rightarrow x = y))$.

(i)

1. $z \neq 0$	Hyp
2. $x \cdot z = 0 \cdot z$	Hyp
3. $0 \cdot z = 0$	Proposition 3.2(l)
4. $x \cdot z = 0$	2, 3 Proposition 3.2(c), MP
5. $x = 0$	1, 4, part(e), MP
6. $\vdash_S z \neq 0 \Rightarrow (x \cdot z = 0 \cdot z \Rightarrow x = 0)$	1–5, deduction theorem
7. $\vdash_S (\forall z)(z \neq 0 \Rightarrow (x \cdot z = 0 \cdot z \Rightarrow x = 0))$	6, Gen

Thus, $\vdash_S \mathscr{B}(0)$.

(ii)

1. $(\forall x)(z \neq 0 \Rightarrow (x \cdot z = y \cdot z \Rightarrow x = y))$	Hyp ($\mathscr{B}(y)$)
2. $z \neq 0$	Hyp
3. $x \cdot z = y' \cdot z$	Hyp
4. $y' \neq 0$	(S3′), Proposition 3.2(b), MP
5. $y' \cdot z \neq 0$	2, 4, part (e), a tautology, MP
6. $x \cdot z \neq 0$	3, 5, (S1′), tautologies, MP
7. $x \neq 0$	6, (S7′), Proposition 3.2(o,n), (S1′), tautologies, MP
8. $(\exists w)(x = w')$	7, part (h), MP
9. $x = b'$	8, rule C
10. $b' \cdot z = y' \cdot z$	3, 9, (A7), MP
11. $b \cdot z + z = y \cdot z + z$	10, Proposition 3.2(m,d), MP
12. $b \cdot z = y \cdot z$	11, Proposition 3.4(d), MP
13. $z \neq 0 \Rightarrow (b \cdot z = y \cdot z \Rightarrow b = y)$	1, rule A4
14. $b \cdot z = y \cdot z \Rightarrow b = y$	2, 13, MP
15. $b = y$	12, 14, MP
16. $b' = y'$	15, (S2′), MP
17. $x = y'$	9, 16, Proposition 3.2(c), MP
18. $\mathscr{B}(y), z \neq 0, x \cdot z = y' \cdot z \vdash_S x = y'$	1–17, Proposition 2.10
19. $\mathscr{B}(y) \vdash_S z \neq 0 \Rightarrow (x \cdot z = y' \cdot z \Rightarrow x = y')$	18, deduction theorem twice
20. $\mathscr{B}(y) \vdash_S (\forall x)(z \neq 0 \Rightarrow (x \cdot z = y' \cdot z \Rightarrow x = y'))$	19, Gen
21. $\vdash_S \mathscr{B}(y) \Rightarrow \mathscr{B}(y')$	20, deduction theorem

Hence, by (i), (ii), Gen, and the induction rule, we obtain $\vdash_S (\forall y)\mathscr{B}(y)$ and then, by Gen and rule A4, we have the desired result.

(j) This is left as an exercise.

PROPOSITION 3.6

(a) Let m and n be any natural numbers.

(i) If $m \neq n$, then $\vdash_S \overline{m} \neq \overline{n}$.
(ii) $\vdash_S \overline{m+n} = \overline{m} + \overline{n}$ and $\vdash_S \overline{m \cdot n} = \overline{m} \cdot \overline{n}$.

(b) Any model for S is infinite.

(c) For any cardinal number $\aleph_\beta$, S has a normal model of cardinality $\aleph_\beta$.

Proof

(a)(i) Assume $m \neq n$. Either $m < n$ or $n < m$. Say, $m < n$.

1. $\overline{m} = \overline{n}$ (Hyp)
2. $\underbrace{0''\ldots'}_{m \text{ times}} = \underbrace{0'''\ldots'}_{n \text{ times}}$ 1 is an abbreviation of 2

3. Apply (S4′) and MP m times in a row. We get $0 = \underbrace{0''\ldots'}_{n-m \text{ times}} \cdot$ Let t be $\overline{n-m-1}$. Since $n > m, n-m-1 \geqslant 0$. Thus, we obtain $0 = t'$.
4. $0 \neq t'$ (S3′)
5. $0 = t' \wedge 0 \neq t'$ 3, 4, conjunction introduction
6. $\overline{m} = \overline{n} \vdash_S 0 = t' \wedge 0 \neq t'$ 1–5
7. $\vdash_S \overline{m} \neq \overline{n}$ 1–6, proof by contradiction

A similar proof holds in the case when $n < m$. (A more rigorous proof can be given by induction in the metalanguage with respect to n.)

(ii) We use induction in the metalanguage. First, $\overline{m+0}$ is $\overline{m}$. Hence, $\vdash_S \overline{m+0} = \overline{m} + \overline{0}$ by (S5′). Now assume $\vdash_S \overline{m+n} = \overline{m} + \overline{n}$. Then $\vdash_S (\overline{m+n})' = \overline{m} + (\overline{n})'$ by (S2′) and (S6′). But $\overline{m+(n+1)}$ is $(\overline{m+n})'$ and $\overline{n+1}$ is $(\overline{n})'$. Hence, $\vdash_S m + (n+1) = \overline{m} + \overline{n+1}$. Thus, $\vdash_S \overline{m+n} = \overline{m} + \overline{n}$. The proof that $\vdash_S \overline{m \cdot n} = \overline{m} \cdot \overline{n}$ is left as an exercise.

(b) By part (a), (i), in a model for S the objects corresponding to the numerals must be distinct. But there are denumberably many numerals.

(c) This follows from Corollary 2.34(c) and the fact that the standard model is an infinite normal model.

An order relation can be introduced by definition in S.

DEFINITIONS

$t < s$ for $(\exists w)(w \neq 0 \wedge w + t = s)$
$t \leqslant s$ for $t < s \vee t = s$
$t > s$ for $s < t$
$t \geqslant s$ for $s \leqslant t$
$t \not< s$ for $\neg(t < s)$, and so on

In the first definition, as usual, we choose w to be the first variable not in t or s.

PROPOSITION 3.7

For any terms t,r,s, the following are theorems.

(a) $t \not< t$
(b) $t < s \Rightarrow (s < r \Rightarrow t < r)$
(c) $t < s \Rightarrow s \not< t$
(d) $t < s \Leftrightarrow t + r < s + r$
(e) $t \leqslant t$
(f) $t \leqslant s \Rightarrow (s \leqslant r \Rightarrow t \leqslant r)$
(g) $t \leqslant s \Leftrightarrow t + r \leqslant s + r$
(h) $t \leqslant s \Rightarrow (s < r \Rightarrow t < r)$
(i) $0 \leqslant t$
(j) $0 < t'$
(k) $t < r \Leftrightarrow t' \leqslant r$
(l) $t \leqslant r \Leftrightarrow t < r'$
(m) $t < t'$
(n) $0 < \overline{1},\ \overline{1} < \overline{2},\ \overline{2} < \overline{3}, \ldots$
(o) $t \neq r \Rightarrow (t < r \vee r < t)$
(p) $t = r \vee t < r \vee r < t$
(q) $t \leqslant r \vee r \leqslant t$
(r) $t + r \geqslant t$
(s) $r \neq 0 \Rightarrow t + r > t$
(t) $r \neq 0 \Rightarrow t \cdot r \geqslant t$
(u) $r \neq 0 \Leftrightarrow r > 0$
(v) $r > 0 \Rightarrow (t > 0 \Rightarrow r \cdot t > 0)$
(w) $r \neq 0 \Rightarrow (t > 1 \Rightarrow t \cdot r > r)$
(x) $r \neq 0 \Rightarrow (t < s \Leftrightarrow t \cdot r < s \cdot r)$
(y) $r \neq 0 \Rightarrow (t \leqslant s \Leftrightarrow t \cdot r \leqslant s \cdot r)$
(z) $t \not< 0$
(z′) $t \leqslant r \wedge r \leqslant t \Rightarrow t = r$

Proof

(a)	1. $t < t$	Hyp
	2. $(\exists w)(w \neq 0 \wedge w + t = t)$	1 is an abbreviation of 2
	3. $b \neq 0 \wedge b + t = t$	2, rule C
	4. $b + t = t$	3, conjunction rule
	5. $t = 0 + t$	Proposition 3.2(f)
	6. $b + t = 0 + t$	3, 4, Proposition 3.2(c), MP
	7. $b = 0$	6, Proposition 3.4(d), MP
	8. $b \neq 0$	3, conjunction elimination
	9. $b = 0 \wedge b \neq 0$	7, 8, conjunction elimination
	10. $0 = 0 \wedge 0 \neq 0$	9, tautology: $B \wedge \neg B \Rightarrow C$, MP
	11. $t < t \vdash_S 0 = 0 \wedge 0 \neq 0$	1–10, Proposition 2.10
	12. $\vdash_S t \not< t$	1–11, proof by contradiction
(b)	1. $t < s$	Hyp
	2. $s < r$	Hyp
	3. $(\exists w)(w \neq 0 \wedge w + t = s)$	1 is an abbreviation of 3
	4. $(\exists v)(v \neq 0 \wedge v + s = r)$	2 is an abbreviation of 4
	5. $b \neq 0 \wedge b + t = s$	3, rule C
	6. $c \neq 0 \wedge c + s = r$	4, rule C
	7. $b + t = s$	5, conjunction elimination
	8. $c + s = r$	6, conjunction elimination
	9. $c + (b + t) = c + s$	7, Proposition 3.2(i), MP
	10. $c + (b + t) = r$	9, 8, Proposition 3.2(c), MP
	11. $(c + b) + t = r$	10, Proposition 3.2(j,c), MP
	12. $b \neq 0$	5, conjunction elimination

13. $c+b\neq 0$	12, Proposition 3.5(d), tautology, MP
14. $c+b\neq 0 \wedge (c+b)+t=r$	13, 11, conjunction introduction
15. $(\exists u)(u\neq 0 \wedge u+t=r)$	14, rule E4
16. $t<r$	Abbreviation of 15
17. $\vdash_S t<s \Rightarrow (s<r \Rightarrow t<r)$	1–15, Proposition 2.10, deduction theorem

Parts (c)–(z′) are left as exercises.

PROPOSITION 3.8

(a) For any natural number $k, \vdash_S x=0 \vee \ldots \vee x=\overline{k} \Leftrightarrow x\leqslant\overline{k}$.
(a′) For any natural number k and any wf $\mathscr{B}$, $\vdash_S \mathscr{B}(0) \wedge \mathscr{B}(\overline{1}) \wedge \ldots \wedge \mathscr{B}(\overline{k}) \Leftrightarrow (\forall x)(x\leqslant\overline{k} \Rightarrow \mathscr{B}(x))$.
(b) For any natural number $k>0, \vdash_S x=0 \vee \ldots \vee x=(\overline{k-1}) \Leftrightarrow x<\overline{k}$
(b′) For any natural number $k>0$ and any wf $\mathscr{B}$, $\vdash_S \mathscr{B}(0) \wedge \mathscr{B}(\overline{1}) \wedge \ldots \wedge \mathscr{B}(\overline{k-1}) \Leftrightarrow (\forall x)(x<\overline{k} \Rightarrow \mathscr{B}(x))$.
(c) $\vdash_S ((\forall x)(x<y \Rightarrow \mathscr{B}(x)) \wedge (\forall x)(x\geqslant y \Rightarrow \mathscr{C}(x))) \Rightarrow (\forall x)(\mathscr{B}(x) \vee \mathscr{C}(x))$

Proof

(a) We prove $\vdash_S x=0 \vee \ldots \vee x=\overline{k} \Leftrightarrow x\leqslant\overline{k}$ by induction in the metalanguage on k. The case for $k=0, \vdash_S x=0 \Leftrightarrow x\leqslant 0$, is obvious from the definitions and Proposition 3.7, Assume as inductive hypothesis $\vdash_S x=0 \vee \ldots \vee x=\overline{k} \Leftrightarrow x\leqslant\overline{k}$. Now assume $x=0 \vee \ldots \vee x=\overline{k} \vee x=\overline{k+1}$. But $\vdash_S x=\overline{k+1} \Rightarrow x\leqslant\overline{k+1}$ and, by the inductive hypothesis, $\vdash_S x=0 \vee \ldots \vee x=\overline{k} \Rightarrow x\leqslant\overline{k}$. Also $\vdash_S x\leqslant\overline{k} \Rightarrow x\leqslant\overline{k+1}$. Thus, $x\leqslant\overline{k+1}$. So, $\vdash_S x=0 \vee \ldots \vee x=\overline{k} \vee x=\overline{k+1} \Rightarrow x\leqslant\overline{k+1}$. Conversely, assume $x\leqslant\overline{k+1}$. Then $x=\overline{k+1} \vee x<\overline{k+1}$. If $x=\overline{k+1}$, then $x=0 \vee \ldots \vee x=\overline{k} \vee x=\overline{k+1}$. If $x<\overline{k+1}$, then since $\overline{k+1}$ is $(\overline{k})'$, we have $x\leqslant\overline{k}$ by Proposition 3.7(l). By the inductive hypothesis, $x=0 \vee \ldots \vee x=\overline{k}$, and, therefore, $x=0 \vee \ldots \vee x=\overline{k} \vee x=\overline{k+1}$. In either case, $x=0 \vee \ldots \vee x=\overline{k} \vee x=\overline{k+1}$. This proves $\vdash_S x\leqslant\overline{k+1} \Rightarrow x=0 \vee \ldots \vee x=\overline{k} \vee x=\overline{k+1}$. From the inductive hypothesis, we have derived $\vdash_S x=0 \vee \ldots \vee x=\overline{k+1} \Leftrightarrow x\leqslant\overline{k+1}$ and this completes the proof. (This proof has been given in an informal manner that we shall generally use from now on. In particular, the deduction theorem, the eliminability of rule C, the replacement theorem, and various derived rules and tautologies will be applied without being explicitly mentioned.)

Parts (a′), (b), and (b′) follow easily from part (a). Part (c) follows almost immediately from Proposition 3.7(o), using obvious tautologies.

There are several stronger forms of the induction principle that we can prove at this point.

PROPOSITION 3.9

(a) *Complete induction.* $\vdash_S (\forall x)((\forall z)(z < x \Rightarrow \mathscr{B}(z)) \Rightarrow \mathscr{B}(x)) \Rightarrow (\forall x)\mathscr{B}(x)$. In ordinary language, consider a property P such that, for any x, if P holds for all natural numbers less than x, then P holds for x also. Then P holds for all natural numbers.

(b) *Least-number principle.* $\vdash_S (\exists x)\mathscr{B}(x) \Rightarrow (\exists y)(\mathscr{B}(y) \wedge (\forall z)(z < y \Rightarrow \neg\mathscr{B}(z)))$. If a property P holds for some natural number, then there is a least number satisfying P.

Proof

(a) Let $\mathscr{C}(x)$ be $(\forall z)(z \leqslant x \Rightarrow \mathscr{B}(z))$.

(i)

1.	$(\forall x)((\forall z)(z < x \Rightarrow \mathscr{B}(z)) \Rightarrow \mathscr{B}(x))$	Hyp
2.	$(\forall z)(z < 0 \Rightarrow \mathscr{B}(z)) \Rightarrow \mathscr{B}(0)$	1, rule A4
3.	$z \not< 0$	Proposition 3.7(y)
4.	$(\forall z)(z < 0 \Rightarrow \mathscr{B}(z))$	3, tautology, Gen
5.	$\mathscr{B}(0)$	2, 4, MP
6.	$(\forall z)(z \leqslant 0 \Rightarrow \mathscr{B}(z))$ i.e., $\mathscr{C}(0)$	5, Proposition 3.8(a′)
7.	$(\forall x)((\forall z)(z < x \Rightarrow \mathscr{B}(z)) \Rightarrow \mathscr{B}(x)) \vdash_S \mathscr{C}(0)$	1–6

(ii)

1.	$(\forall x)((\forall z)(z < x \Rightarrow \mathscr{B}(z)) \Rightarrow \mathscr{B}(x))$	Hyp
2.	$\mathscr{C}(x)$, i.e., $(\forall z)(z \leqslant x \Rightarrow \mathscr{B}(z))$	Hyp
3.	$(\forall z)(z < x' \Rightarrow \mathscr{B}(z))$	2, Proposition 3.7(ℓ)
4.	$(\forall z)(z < x' \Rightarrow \mathscr{B}(z)) \Rightarrow \mathscr{B}(x')$	1, rule A4
5.	$\mathscr{B}(x')$	3, 4, MP
6.	$z \leqslant x' \Rightarrow z < x' \vee z = x'$	Definition, tautology
7.	$z < x' \Rightarrow \mathscr{B}(z)$	3, rule A4
8.	$z = x' \Rightarrow \mathscr{B}(z)$	5, axiom (A7), Proposition 2.23(b), tautologies
9.	$(\forall z)(z \leqslant x' \Rightarrow \mathscr{B}(z))$ i.e., $\mathscr{C}(x')$	6, 7, 8, Tautology, Gen
10.	$(\forall x)((\forall z)(z < x \Rightarrow \mathscr{B}(z)) \Rightarrow \mathscr{B}(x)) \vdash_S (\forall x)(\mathscr{C}(x) \Rightarrow \mathscr{C}(x'))$	1–9, deduction theorem, Gen

By (i), (ii) and the induction rule, we obtain $\mathscr{D} \vdash_S (\forall x)\mathscr{C}(x)$, that is, $\mathscr{D} \vdash_S (\forall x)(\forall z)(z \leqslant x \Rightarrow \mathscr{B}(z))$, where $\mathscr{D}$ is $(\forall x)((\forall z)(z < x \Rightarrow \mathscr{B}(z)) \Rightarrow \mathscr{B}(x))$. Hence, by rule A4 twice, $\mathscr{D} \vdash_S x \leqslant x \Rightarrow \mathscr{B}(x)$. But $\vdash_S x \leqslant x$. So, $\mathscr{D} \vdash_S \mathscr{B}(x)$, and, by Gen and the deduction theorem, $\vdash_S \mathscr{D} \Rightarrow (\forall x)\mathscr{B}(x)$.

(b)

1.	$\neg(\exists y)(\mathscr{B}(y) \wedge (\forall z)(z < y \Rightarrow \neg\mathscr{B}(z)))$	Hyp
2.	$(\forall y)\neg(\mathscr{B}(y) \wedge (\forall z)(z < y \Rightarrow \neg\mathscr{B}(z)))$	1, derived rule for negation
3.	$(\forall y)((\forall z)(z < y \Rightarrow \neg\mathscr{B}(z)) \Rightarrow \neg\mathscr{B}(y))$	2, tautology, replacement
4.	$(\forall y)\neg\mathscr{B}(y)$	3, part (a) with $\neg\mathscr{B}$ instead of $\mathscr{B}$
5.	$\neg(\exists y)\mathscr{B}(y)$	4, derived rule for negation

6. $\neg(\exists x)\mathscr{B}(x)$ — 5, change of bound variable
7. $\vdash_S \neg(\exists y)(\mathscr{B}(y) \wedge (\forall z)(z < y \Rightarrow \neg\mathscr{B}(z))) \Rightarrow \neg(\exists x)\mathscr{B}(x)$ — 1–6, deduction theorem
8. $\vdash_S (\exists x)\mathscr{B}(x) \Rightarrow (\exists y)(\mathscr{B}(y) \wedge (\forall z)(z < y \Rightarrow \neg\mathscr{B}(z)))$ — 7, derived rule

Exercise

3.1 (*Method of infinite descent*)
Prove $\vdash_S (\forall x)(\mathscr{B}(x) \Rightarrow (\exists y)(y < x \wedge \mathscr{B}(y))) \Rightarrow (\forall x)\neg\mathscr{B}(x)$

Another important notion in number theory is divisibility, which we now define.

DEFINITION $t|s$ for $(\exists z)(s = t \cdot z)$. (Here, z is the first variable not in t or s.)

PROPOSITION 3.10

The following wfs are theorems for any terms t, s, r.

(a) $t|t$
(b) $\overline{1}|t$
(c) $t|0$
(d) $t|s \wedge s|r \Rightarrow t|r$
(e) $s \neq 0 \wedge t|s \Rightarrow t \leqslant s$
(f) $t|s \wedge s|t \Rightarrow s = t$
(g) $t|s \Rightarrow t|(r \cdot s)$
(h) $t|s \wedge t|r \Rightarrow t|(s + r)$

Proof

(a) $t = t \cdot \overline{1}$. Hence, $t|t$.
(b) $t = \overline{1} \cdot t$. Hence, $\overline{1}|t$.
(c) $0 = t \cdot 0$. Hence, $t|0$.
(d) If $s = t \cdot z$ and $r = s \cdot w$, then $r = t \cdot (z \cdot w)$.
(e) If $s \neq 0$ and $t|s$, then $s = t \cdot z$ for some z. If $z = 0$, then $s = 0$. Hence, $z \neq 0$. So, $z = u'$ for some u. Then $s = t \cdot (u') = t \cdot u + t \geqslant t$.
(f)–(h) These proofs are left as exercises.

Exercises

3.2 Prove $\vdash_S t|\overline{1} \Rightarrow t = \overline{1}$.
3.3 Prove $\vdash_S (t|s \wedge t|s') \Rightarrow t = 1$.

It will be useful for later purposes to prove the existence of a unique quotient and remainder upon division of one number x by another nonzero number y.

PROPOSITION 3.11

$\vdash_S y \neq 0 \Rightarrow (\exists u)\ (\exists v)[x = y \cdot u + v \wedge v < y \wedge (\forall u_1)\ (\forall v_1)\ ((x = y \cdot u_1 + v_1 \wedge v_1 < y) \Rightarrow u = u_1 \wedge v = v_1)]$

Proof

Let $\mathscr{B}(x)$ be $y \neq 0 \Rightarrow (\exists u)(\exists v)(x = y \cdot u + v \wedge v < y)$.

(i)	1. $y \neq 0$	Hyp
	2. $0 = y \cdot 0 + 0$	(S5′), (S7′)
	3. $0 < y$	1, Proposition 3.7(t)
	4. $0 = y \cdot 0 + 0 \wedge 0 < y$	2, 3, conjunction rule
	5. $(\exists u)(\exists v)(0 = y \cdot u + v \wedge v < y)$	4, rule E4 twice
	6. $y \neq 0 \Rightarrow (\exists u)(\exists v)(0 = y \cdot u + v \wedge v < y)$	1–5, deduction theorem
(ii)	1. $\mathscr{B}(x)$ i.e., $y \neq 0 \Rightarrow (\exists u)(\exists v)(x = y \cdot u + v \wedge v < y)$	Hyp
	2. $y \neq 0$	Hyp
	3. $(\exists u)(\exists v)(x = y \cdot u + v \wedge v < y)$	1, 2, MP
	4. $x = y \cdot a + b \wedge b < y$	3, rule C twice
	5. $b < y$	4, conjunction elimination
	6. $b' \leqslant y$	5, Proposition 3.7(k)
	7. $b' < y \vee b' = y$	6, definition
	8. $b' < y \Rightarrow (x' = y \cdot a + b' \wedge b' < y)$	4, (S6′), derived rules
	9. $b' < y \Rightarrow (\exists u)(\exists v)(x' = y \cdot u + v \wedge v < y)$	8, rule E4, deduction theorem
	10. $b' = y \Rightarrow x' = y \cdot a + y \cdot \bar{1}$	4, (S6′), Proposition 3.5(b)
	11. $b' = y \Rightarrow (x' = y \cdot (a + \bar{1}) + 0 \wedge 0 < y)$	10, Proposition 3.4, 2, Proposition 3.7(t), (S5′)
	12. $b' = y \Rightarrow (\exists u)(\exists v)(x' = y \cdot u + v \wedge v < y)$	11, rule E4 twice, deduction theorem
	13. $(\exists u)(\exists v)(x' = y \cdot u + v \wedge v < y)$	7, 9, 12, disjunction elimination
	14. $\mathscr{B}(x) \Rightarrow (y \neq 0 \Rightarrow (\exists u)(\exists v)(x' = y \cdot u + v \wedge v < y))$ i.e., $\mathscr{B}(x) \Rightarrow \mathscr{B}(x')$	1–13, deduction theorem

By (i), (ii), Gen and the induction rule, $\vdash_S (\forall x)\mathscr{B}(x)$. This establishes the existence of a quotient u and a remainder v. To prove uniqueness, proceed as follows. Assume $y \neq 0$. Assume $x = y \cdot u + v \wedge v < y$ and $x = y \cdot u_1 +$

$v_1 \wedge v_1 < y$. Now, $u = u_1$ or $u < u_1$ or $u_1 < u$. If $u = u_1$, then $v = v_1$ by Proposition 3.4(d). If $u < u_1$, then $u_1 = u + w$ for some $w \neq 0$. Then $y \cdot u + v = y \cdot (u + w) + v_1 = y \cdot u + y \cdot w + v_1$. Hence, $v = y \cdot w + v_1$. Since $w \neq 0, y \cdot w \geqslant y$. So, $v = y \cdot w + v_1 \geqslant y$, contradicting $v < y$. Hence, $u \not< u_1$. Similarly, $u_1 \not< u$. Thus, $u = u_1$. Since $y \cdot u + v = x = y \cdot u_1 + v_1$, it follows that $v = v_1$.

From this point on, one can generally translate into S and prove the results from any text on elementary number theory. There are certain number-theoretic functions, such as x^y and $x!$, that we have to be able to define in S, and this we shall do later in this chapter. Some standard results of number theory, such as Dirichlet's theorem, are proved with the aid of the theory of complex variables, and it is often not known whether elementary proofs (or proofs in S) can be given for such theorems. The statement of some results in number theory involves non-elementary concepts, such as the logarithmic function, and, except in special cases, cannot even be formulated in S. More information about the strength and expressive powers of S will be revealed later. For example, it will be shown that there are closed wfs that are neither provable nor disprovable in S, if S is consistent; hence there is a wf that is true under the standard interpretation but is not provable in S. We also will see that this incompleteness of S cannot be attributed to omission of some essential axiom but has deeper underlying causes that apply to other theories as well.

Exercises

3.4 Show that the induction principle (S9) is independent of the other axioms of S.

3.5^D

(a) Show that there exist non-standard models for S of any cardinality $\aleph_\alpha$.

(b) Ehrenfeucht (1958) has shown the existence of at least $2^{\aleph_0}$ mutually non-isomorphic models of cardinality $\aleph_\alpha$. Prove the special case that there are $2^{\aleph_0}$ mutually non-isomorphic denumerable models of S.

3.6^D Give a standard mathematical proof of the categoricity of Peano's postulates, in the sense that any two 'models' are isomorphic. Explain why this proof does not apply to the first-order theory S.

3.7^D (Presburger, 1929) If we eliminate from S the function letter f_2^2 for multiplication and the axioms (S7) and (S8), show that the new system S_+ is complete and decidable (in the sense of Chapter 1, p. 34).

3.8

(a) Show that, for every closed term t of S, we can find a natural number n such that $\vdash_S t = \overline{n}$.

(b) Show that every closed atomic wf $t = s$ of S is *decidable* – that is, either $\vdash_S t = s$ or $\vdash_S t \neq s$.

(c) Show that every closed wf of S without quantifiers is decidable.

3.2 NUMBER-THEORETIC FUNCTIONS AND RELATIONS

A *number-theoretic function* is a function whose arguments and values are natural numbers. Addition and multiplication are familiar examples of number-theoretic functions of two arguments. By a *number-theoretic relation* we mean a relation whose arguments are natural numbers. For example, $=$ and $<$ are binary number-theoretic relations, and the expression $x+y<z$ determines a number-theoretic relation of three arguments.† Number-theoretic functions and relations are intuitive and are not bound up with any formal system.

Let K be any theory in the language $\mathscr{L}_A$ of arithmetic. We say that a number-theoretic relation R of n arguments is *expressible* in K if and only if there is a wf $\mathscr{B}(x_1,\ldots,x_n)$ of K with the free variables $x_1,\ldots,x_n$ such that, for any natural numbers $k_1,\ldots,k_n$, the following hold:

1. If $\mathrm{R}(k_1,\ldots,k_n)$ is true, then $\vdash_{\mathrm{K}} \mathscr{B}(\overline{k}_1,\ldots,\overline{k}_n)$.
2. If $\mathrm{R}(k_1,\ldots,k_n)$ is false, then $\vdash_{\mathrm{K}} \neg\mathscr{B}(\overline{k}_1,\ldots,\overline{k}_n)$.

For example, the number-theoretic relation of identity is expressed in S by the wf $x_1 = x_2$. In fact, if $k_1 = k_2$, then $\overline{k}_1$ is the same term as $\overline{k}_2$ and so, by Proposition 3.2(a), $\vdash_{\mathrm{S}} \overline{k}_1 = \overline{k}_2$. Moreover, if $k_1 \neq k_2$, then, by Proposition 3.6(a), $\vdash_{\mathrm{S}} \overline{k}_1 \neq \overline{k}_2$.

Likewise, the relation 'less than' is expressed in S by the wf $x_1 < x_2$. Recall that $x_1 < x_2$ is $(\exists x_3)(x_3 \neq 0 \wedge x_3 + x_1 = x_2)$. If $k_1 < k_2$, then there is some non-zero number n such that $k_2 = n + k_1$. Now, by Proposition 3.6(a)(ii), $\vdash_{\mathrm{S}} \overline{k}_2 = \overline{n} + \overline{k}_1$. Also, by (S3′), since $n \neq 0$, $\vdash_{\mathrm{S}} \overline{n} \neq 0$. Hence, by rule E4, one can prove in S the wf $(\exists w)(w \neq 0 \wedge w + \overline{k}_1 = \overline{k}_2)$; that is, $\vdash_{\mathrm{S}} \overline{k}_1 < \overline{k}_2$. On the other hand, if $k_1 \not< k_2$, then $k_2 < k_1$ or $k_2 = k_1$. If $k_2 < k_1$, then, as we have just seen, $\vdash_{\mathrm{S}} \overline{k}_2 < \overline{k}_1$. If $k_2 = k_1$, then $\vdash_{\mathrm{S}} \overline{k}_2 = \overline{k}_1$. In either case, $\vdash_{\mathrm{S}} \overline{k}_2 \leqslant \overline{k}_1$ and then, by Proposition 3.7(a,c), $\vdash_{\mathrm{S}} \overline{k}_1 \not< \overline{k}_2$.

Observe that, if a relation is expressible in a theory K, then it is expressible in any extension of K.

Exercises

3.9 Show that the negation, disjunction, and conjunction of relations that are expressible in K are also expressible in K.

3.10 Show that the relation $x + y = z$ is expressible in S.

†We follow the custom of regarding a number-theoretic property, such as the property of being even, as a 'relation' of one argument.

Let K be any theory with equality in the language $\mathscr{L}_A$ of arithmetic. A number-theoretic function f of n arguments is said to be *representable* in K if and only if there is a wf $\mathscr{B}(x_1,\ldots,x_n,y)$ of K with the free variables $x_1,\ldots,x_n,y$ such that, for any natural numbers $k_1,\ldots,k_n,m$, the following hold:

1. If $f(k_1,\ldots,k_n)=m$, then $\vdash_K \mathscr{B}(\bar{k}_1,\ldots,\bar{k}_n,\bar{m})$.
2. $\vdash_K (\exists_1 y)\mathscr{B}(\bar{k}_1,\ldots,\bar{k}_n,y)$.

If, in this definition, we replace condition 2 by

2′. $\vdash_K (\exists_1 y)\mathscr{B}(x_1,\ldots,x_n,y)$

then the function f is said to be *strongly representable* in K. Notice that 2′ implies 2, by Gen and rule A4. Hence, strong representability implies representability. The converse is also true, as we now prove.

PROPOSITION 3.12 (V.H. DYSON)

If $f(x_1,\ldots,x_n)$ is representable in K, then it is strongly representable in K.

Proof

Assume f representable in K by a wf $\mathscr{B}(x_1,\ldots,x_n,y)$. Let us show that f is strongly representable in K by the following wf $\mathscr{C}(x_1,\ldots,x_n,y)$:

$$([(\exists_1 y)\mathscr{B}(x_1,\ldots,x_n,y)]\wedge\mathscr{B}(x_1,\ldots,x_n,y))\vee(\neg[(\exists_1 y)\mathscr{B}(x_1,\ldots,x_n,y)]\wedge y=0)$$

1. Assume $f(k_1,\ldots,k_n)=m$. Then $\vdash_K \mathscr{B}(\bar{k}_1,\ldots,\bar{k}_n,\bar{m})$ and $\vdash_K (\exists_1 y)\mathscr{B}(\bar{k}_1,\ldots,\bar{k}_n,y)$. So, by conjunction introduction and disjunction introduction, we get $\vdash_K \mathscr{C}(\bar{k}_1,\ldots,\bar{k}_n,\bar{m})$.

2′. We must show $\vdash_K (\exists_1 y)\mathscr{C}(x_1,\ldots,x_n,y)$.

Case 1. Take $(\exists_1 y)\mathscr{B}(x_1,\ldots,x_n,y)$ as hypothesis. (i) It is easy, using rule C, to obtain $\mathscr{B}(x_1,\ldots,x_n,b)$ from our hypothesis, where b is a new individual constant. Together with our hypothesis and conjunction and disjunction introduction, this yields $\mathscr{C}(x_1,\ldots,x_n,b)$ and then, by rule E4, $(\exists y)\mathscr{C}(x_1,\ldots,x_n,y)$. (ii) Assume $\mathscr{C}(x_1,\ldots,x_n,u)\wedge\mathscr{C}(x_1,\ldots,x_n,v)$. From $\mathscr{C}(x_1,\ldots,x_n,u)$ and our hypothesis, we obtain $\mathscr{B}(x_1,\ldots,x_n,u)$, and, from $\mathscr{C}(x_1,\ldots,x_n,v)$ and our hypothesis, we obtain $\mathscr{B}(x_1,\ldots,x_n,v)$. Now, from $\mathscr{B}(x_1,\ldots,x_n,u)$ and $\mathscr{B}(x_1,\ldots,x_n,v)$ and our hypothesis, we get $u=v$. The deduction theorem yields $\mathscr{C}(x_1,\ldots,x_n,u)\wedge\mathscr{C}(x_1,\ldots,x_n,v)\Rightarrow u=v$. From (i) and (ii), $(\exists_1 y)\mathscr{C}(x_1,\ldots,x_n,y)$. Thus, we have proved $\vdash_K (\exists_1 y)\ \mathscr{B}(x_1,\ldots,x_n,y)\Rightarrow(\exists_1 y)\mathscr{C}(x_1,\ldots,x_n,y)$.

Case 2. Take $\neg(\exists_1 y)\mathscr{B}(x_1,\ldots,x_n,y)$ as hypothesis. (i) Our hypothesis, together with the theorem $0=0$, yields, by conjunction introduction, $\neg(\exists_1 y)\mathscr{B}(x_1,\ldots,x_n,y)\wedge 0=0$. By disjunction introduction, $\mathscr{C}(x_1,\ldots,x_n,0)$,

and, by rule E4, $(\exists y)\mathscr{C}(x_1,\ldots,x_n,y)$. (ii) Assume $\mathscr{C}(x_1,\ldots,x_n,u)\wedge \mathscr{C}(x_1,\ldots,x_n,v)$. From $\mathscr{C}(x_1,\ldots,x_n,u)$ and our hypothesis, it follows easily that $u=0$. Likewise, from $\mathscr{C}(x_1,\ldots,x_n,v)$ and our hypothesis, $v=0$. Hence, $u=v$. By the deduction theorem, $\mathscr{C}(x_1,\ldots,x_n,u)\wedge\mathscr{C}(x_1,\ldots,x_n,v)\Rightarrow u=v$. From (i) and (ii), $(\exists_1 y)\mathscr{C}(x_1,\ldots,x_n,y)$. Thus we have proved $\vdash_{\mathrm{K}} \neg(\exists_1 y)\mathscr{B}(x_1,\ldots,x_n,y)\Rightarrow(\exists_1 y)\mathscr{C}(x_1,\ldots,x_n,y)$

By case 1 and case 2 and an instance of the tautology $[(D\Rightarrow E)\wedge(\neg D\Rightarrow E)]\Rightarrow E$, We can obtain $\vdash_{\mathrm{K}}(\exists_1 y)\mathscr{C}(x_1,\ldots,x_n,y)$.

Since we have proved them to be equivalent, from now on we shall use representability and strong representability interchangeably.

Observe that a function representable in K is representable in any extension of K.

Examples

In these examples, let K be any theory with equality in the language $\mathscr{L}_A$.

1. The zero function, $Z(x)=0$, is representable in K by the wf $x_1=x_1\wedge y=0$. For any k and m, if $Z(k)=m$, then $m=0$ and $\vdash_{\mathrm{K}}\bar{k}=\bar{k}\wedge 0=0$; that is, condition 1 holds. Also, it is easy to show that $\vdash_{\mathrm{K}}(\exists_1 y)(x_1=x_1\wedge y=0)$. Thus, condition 2′ holds.
2. The successor function, $N(x)=x+1$, is representable in K by the wf $y=x_1'$. For any k and m, if $N(k)=m$, then $m=k+1$; hence, $\bar{m}$ is $\bar{k}'$. Then $\vdash_{\mathrm{K}}\bar{m}=\bar{k}'$. It is easy to verify that $\vdash_{\mathrm{K}}(\exists_1 y)(y=x_1')$.
3. The projection function, $U_j^n(x_1,\ldots,x_n)=x_j$, is representable in K by $x_1=x_1\wedge x_2=x_2\wedge\ldots\wedge x_n=x_n\wedge y=x_j$. If $U_j^n(k_1,\ldots,k_n)=m$, then $m=k_j$. Hence, $\vdash_{\mathrm{K}}\bar{k}_1=\bar{k}_1\wedge\bar{k}_2=\bar{k}_2\wedge\ldots\wedge\bar{k}_n=\bar{k}_n\wedge\bar{m}=\bar{k}_j$. Thus, condition 1 holds. Also, $\vdash_{\mathrm{K}}(\exists_1 y)(x_1=x_1\wedge x_2=x_2\wedge\ldots\wedge x_n=x_n\wedge y=x_j)$, that is, condition 2′ holds.
4. Assume that the functions $g(x_1,\ldots,x_m), h_1(x_1,\ldots,x_n),\ldots, h_m(x_1,\ldots,x_n)$ are strongly representable in the theory with equality K by the wfs $\mathscr{C}(x_1,\ldots,x_m,z), \mathscr{B}_1(x_1,\ldots,x_n,y_1),\ldots,\mathscr{B}_m(x_1,\ldots,x_n,y_m)$, respectively. Define a new function f by the equation

$$f(x_1,\ldots x_n)=g(h_1(x_1,\ldots,x_n),\ldots,h_m(x_1,\ldots,x_n))$$

f is said to be obtained from $g,h_1,\ldots,h_m$ by *substitution*. Then f is also strongly representable in K by the following wf $\mathscr{D}(x_1,\ldots,x_n,z)$:

$$(\exists y_1)\ldots(\exists y_m)(\mathscr{B}_1(x_1,\ldots,x_n,y_1)\wedge\ldots\wedge\mathscr{B}_m(x_1,\ldots,x_n,y_m)\wedge\mathscr{C}(y_1,\ldots,y_m,z))$$

To prove condition 1, let $f(k_1,\ldots,k_n)=p$. Let $h_j(k_1,\ldots,k_n)=r_j$ for $1\leqslant j\leqslant m$; then $g(r_1,\ldots,r_m)=p$. Since $\mathscr{C},\mathscr{B}_1,\ldots,\mathscr{B}_m$ represent $g,h_1,\ldots,h_m$, we have $\vdash_{\mathrm{K}}\mathscr{B}_j(\bar{k}_1,\ldots,\bar{k}_n,\bar{r}_j)$ for $1\leqslant j\leqslant m$ and $\vdash_{\mathrm{K}}\mathscr{C}(\bar{r}_1,\ldots,\bar{r}_m,\bar{p})$. So by conjunction introduction, $\vdash_{\mathrm{K}}\mathscr{B}_1(\bar{k}_1,\ldots,\bar{k}_n,\bar{r}_1)\wedge\ldots\wedge\mathscr{B}_m(\bar{k}_1,\ldots,\bar{k}_n,\bar{r}_m)\wedge\mathscr{C}(\bar{r}_1\ldots\bar{r}_m,\bar{p})$. Hence, by rule $E4$, $\vdash_{\mathrm{K}}\mathscr{D}(\bar{k}_1,\ldots,\bar{k}_n,\bar{p})$. Thus, condition 1 holds. Now we shall prove condition 2′. Assume $\mathscr{D}(x_1,\ldots,x_n,u)\wedge\mathscr{D}(x_1,\ldots,x_n,v)$, that is

$$(\Delta)(\exists y_1)\ldots(\exists y_m)(\mathscr{B}_1(x_1,\ldots,x_n,y_1)\wedge\ldots\wedge\mathscr{B}_m(x_1,\ldots,x_n,y_m)\wedge\mathscr{C}(y_1,\ldots,y_m,u))$$

and

$$(\square)(\exists y_1)\ldots(\exists y_m)(\mathscr{B}_1(x_1,\ldots,x_n,y_1)\wedge\ldots\wedge\mathscr{B}_m(x_1,\ldots,x_n,y_m)\wedge\mathscr{C}(y_1,\ldots,y_m,v))$$

By(Δ), using rule C m times,

$$\mathscr{B}_1(x_1,\ldots,x_n,b_1)\wedge\ldots\wedge\mathscr{B}_m(x_1,\ldots,x_n,b_m)\wedge\mathscr{C}(b_1,\ldots,b_m,u)$$

By($\square$) using rule C again,

$$\mathscr{B}_1(x_1,\ldots,x_n,c_1)\wedge\ldots\wedge\mathscr{B}_m(x_1,\ldots,x_n,c_m)\wedge\mathscr{C}(c_1,\ldots,c_m,v)$$

Since $\vdash_K(\exists_1 y_j)\mathscr{B}_j(x_1,\ldots,x_n,y_j)$, we obtain from $\mathscr{B}_j(x_1,\ldots,x_n,b_j)$ and $B_j(x_1,\ldots,x_n,c_j)$, that $b_j=c_j$. From $\mathscr{C}(b_1,\ldots,b_m,u)$ and $b_1=c_1,\ldots,b_m=c_m$, we have $\mathscr{C}(c_1,\ldots,c_m,u)$. This, with $\vdash_K(\exists_1 z)\mathscr{C}(x_1,\ldots,x_n,z)$ and $\mathscr{C}(c_1,\ldots,c_m,v)$ yields $u=v$. Thus, we have shown $\vdash_K\mathscr{D}(x_1,\ldots,x_n,u)\wedge\mathscr{D}(x_1,\ldots,x_n,v)\Rightarrow u=v$. It is easy to show that $\vdash_K(\exists z)\mathscr{D}(x_1,\ldots,x_n,z)$. Hence, $\vdash_K(\exists_1 z)\mathscr{D}(x_1,\ldots,x_n,z)$.

Exercises

3.11 Let K be a theory with equality in the language $\mathscr{L}_A$. Show that the following functions are representable in K.

(a) $Z_n(x_1,\ldots,x_n)=0$ [*Hint* : $Z_n(x_1,\ldots,x_n)=Z(U_1^n(x_1,\ldots,x_n))$.]

(b) $C_k^n(x_1,\ldots,x_n)=k$, where k is a fixed natural number. [*Hint:* Use mathematical induction in the metalanguage with respect to k.]

3.12 Prove that addition and multiplication are representable in S.

If R is a relation of n arguments, then the characteristic function C_R is defined as follows:

$$C_R(x_1,\ldots,x_n)=\begin{cases}0 & \text{if } R(x_1,\ldots x_n) \text{ is true}\\ 1 & \text{if } R(x_1,\ldots,x_n) \text{ is false}\end{cases}$$

PROPOSITION 3.13

Let K be a theory with equality in the language $\mathscr{L}_A$ such that $\vdash_K 0\neq\bar{1}$. Then a number-theoretic relation R is expressible in K if and only if C_R is representable in K.

Proof

If R is expressible in K by a wf $\mathscr{B}(x_1,\ldots,x_n)$, it is easy to verify that C_R is representable in K by the wf $(\mathscr{B}(x_1,\ldots,x_n)\wedge y=0)\vee(\neg\mathscr{B}(x_1,\ldots,x_n)\wedge y=\bar{1})$. Conversely, if C_R is representable in K by a wf $\mathscr{C}(x_1,\ldots,x_n,y)$,

then, using the assumption that $\vdash_K 0 \neq \bar{1}$, we can easily show that R is expressible in K by the wf $\mathscr{C}(x_1, \ldots, x_n, 0)$.

Exercises

3.13 The *graph* of a function $f(x_1, \ldots, x_n)$ is the relation $f(x_1, \ldots, x_n) = x_{n+1}$. Show that $f(x_1, \ldots, x_n)$ is representable in S if and only if its graph is expressible in S.

3.14 If Q and R are relations of n arguments, prove that $C_{\text{not-R}} = 1 - C_R$, $C_{(Q \text{ or } R)} = C_Q \cdot C_R$, and $C_{(Q \text{ and } R)} = C_Q + C_R - C_Q \cdot C_R$.

3.15 Show that $f(x_1, \ldots, x_n)$ is representable in a theory with equality K in the language $\mathscr{L}_A$ if and only if there is a wf $\mathscr{B}(x_1, \ldots, x_n, y)$ such that, for any $k_1, \ldots, k_n, m$, if $f(k_1, \ldots, k_n) = m$, then $\vdash_K (\forall y)(\mathscr{B}(\bar{k}_1, \ldots, \bar{k}_n, y) \Leftrightarrow y = \bar{m})$.

3.3 PRIMITIVE RECURSIVE AND RECURSIVE FUNCTIONS

The study of representability of functions in S leads to a class of number-theoretic functions that turn out to be of great importance in mathematical logic and computer science.

DEFINITION

1. The following functions are called *initial functions*.
   (I) The *zero function*, $Z(x) = 0$ for all x.
   (II) The *successor function*, $N(x) = x + 1$ for all x.
   (III) The *projection functions*, $U_i^n(x_1, \ldots, x_n) = x_i$ for all $x_1, \ldots, x_n$.
2. The following are rules for obtaining new functions from given functions.
   (IV) *Substitution*:

$$f(x_1, \ldots, x_n) = g(h_1(x_1, \ldots, x_n), \ldots, h_m(x_1, \ldots, x_n))$$

f is said to be obtained by substitution from the functions

$$g(y_1, \ldots, y_m), h_1(x_1, \ldots, x_n), \ldots, h_m(x_1, \ldots, x_n)$$

(V) *Recursion:*

$$f(x_1, \ldots, x_n, 0) = g(x_1, \ldots, x_n)$$
$$f(x_1, \ldots, x_n, y + 1) = h(x_1, \ldots, x_n, y, f(x_1, \ldots, x_n, y))$$

Here, we allow $n = 0$, in which case we have

$$f(0) = k \quad \text{where } k \text{ is a fixed natural number}$$
$$f(y + 1) = h(y, f(y))$$

We shall say that f is obtained from g and h (or, in the case $n = 0$, from h alone) by recursion. The *parameters* of the recursion are $x_1, \ldots, x_n$. Notice that f is well defined: $f(x_1, \ldots, x_n, 0)$ is given by the first equation, and if we already know $f(x_1, \ldots, x_n, y)$, then we can obtain $f(x_1, \ldots, x_n, y+1)$ by the second equation.

(VI) *Restricted μ-Operator.* Assume that $g(x_1, \ldots, x_n, y)$ is a function such that for any $x_1, \ldots, x_n$ there is at least one y such that $g(x_1, \ldots, x_n, y) = 0$. We denote by $\mu y(g(x_1, \ldots, x_n, y) = 0)$ the least number y such that $g(x_1, \ldots, x_n, y) = 0$. In general, for any relation $R(x_1, \ldots, x_n, y)$, we denote by $\mu y R(x_1, \ldots, x_n, y)$ the least y such that $R(x_1, \ldots, x_n, y)$ is true, if there is any y at all such that $R(x_1, \ldots, x_n, y)$ holds. Let $f(x_1, \ldots, x_n) = \mu y(g(x_1, \ldots, x_n, y) = 0)$. Then f is said to be obtained from g by means of the restricted μ-operator if the given assumption about g holds, namely, for any $x_1, \ldots, x_n$, there is at least one y such that $g(x_1, \ldots, x_n, y) = 0$.

3. A function f is said to be *primitive recursive* if and only if it can be obtained from the initial functions by any finite number of substitutions (IV) and recursions (V) – that is, if there is a finite sequence of functions $f_0, \ldots, f_n$ such that $f_n = f$ and, for $0 \leqslant i \leqslant n$, either f_i is an initial function or f_i comes from preceding functions in the sequence by an application of rule (IV) or rule (V).
4. A function f is said to be *recursive* if and only if it can be obtained from the initial functions by any finite number of applications of substitution (IV), recursion (V) and the restricted μ-operator (VI). This differs from the definition above of primitive recursive functions only in the addition of possible applications of the restricted μ -operator. Hence, every primitive recursive function is recursive. We shall see later that the converse is false.

We shall show that the class of recursive functions is identical with the class of functions representable in S. (In the literature, the phrase 'general recursive' is sometimes used instead of 'recursive'.)

First, let us prove that we can add 'dummy variables' to and also permute and identify variables in any primitive recursive or recursive function, obtaining a function of the same type.

PROPOSITION 3.14

Let $g(y_1, \ldots, y_k)$ be primitive recursive (or recursive). Let $x_1, \ldots x_n$ be distinct variables and, for $1 \leqslant i \leqslant k$, let z_i be one of $x_1, \ldots, x_n$. Then the function f such that $f(x_1, \ldots, x_n) = g(z_1, \ldots, z_k)$ is primitive recursive (or recursive).

Proof

Let $z_i = x_{j_i}$, where $1 \leqslant j_i \leqslant n$. Then $z_i = U^n_{j_i}(x_1, \ldots, x_n)$. Thus,

$$f(x_1, \ldots, x_n) = g(U^n_{j_1}(x_1, \ldots, x_n), \ldots, U^n_{j_k}(x_1, \ldots, x_n))$$

and therefore f is primitive recursive (or recursive), since it arises from g, $U^n_{j_1}, \ldots, U^n_{j_k}$ by substitution.

Examples

1. *Adding dummy variables.* If $g(x_1, x_3)$ is primitive recursive and if $f(x_1, x_2, x_3) = g(x_1, x_3)$, then $f(x_1, x_2, x_3)$ is also primitive recursive. In Proposition 3.14, let $z_1 = x_1$ and $z_2 = x_3$. The new variable x_2 is called a 'dummy variable' since its value has no influence on the value of $f(x_1, x_2, x_3)$.
2. *Permuting variables.* If $g(x_1, x_2, x_3)$ is primitive recursive and if $f(x_1, x_2, x_3) = g(x_3, x_1, x_2)$, then $f(x_1, x_2, x_3)$ is also primitive recursive. In Proposition 3.14, let $z_1 = x_3, z_2 = x_1$ and $z_3 = x_2$.
3. *Identifying variables.* If $g(x_1, x_2, x_3)$ is primitive recursive and if $f(x_1, x_2) = g(x_1, x_2, x_1)$, then $f(x_1, x_2)$ is primitive recursive. In Proposition 3.14, let $n = 2$ and $z_1 = x_1, z_2 = x_2$ and $z_3 = x_1$.

COROLLARY 3.15

(a) The zero function $Z_n(x_1, \ldots, x_n) = 0$ is primitive recursive.
(b) The constant function $C^n_k(x_1, \ldots, x_n) = k$, where k is some fixed natural number, is primitive recursive.
(c) The substitution rule (IV) can be extended to the case where each h_i may be a function of some but not necessarily all of the variables. Likewise, in the recursion rule (V), the function g may not involve all of the variables $x_1, \ldots, x_n, y$, or $f(x_1, \ldots, x_n, y)$ and h may not involve all of the variables $x_1, \ldots, x_n, y$, or $f(x_1, \ldots, x_n, y)$.

Proof

(a) In Proposition 3.14, let g be the zero function Z; then $k = 1$. Take z_1 to be x_1.

(b) Use mathematical induction. For $k = 0$, this is part (a). Assume C^n_k primitve recursive. Then $C^n_{k+1}(x_1, \ldots, x_n)$ is primitive recursive by the substitution $C^n_{k+1}(x_1, \ldots, x_n) = N(C^n_k(x_1, \ldots, x_n))$.

(c) By Proposition 3.14, any variables among $x_1, \ldots, x_n$ not present in a function can be added as dummy variables. For example, if $h(x_1, x_3)$ is primitive recursive, then $h^*(x_1, x_2, x_3) = h(x_1, x_3) = h(U^3_1(x_1, x_2, x_3), U^3_3(x_1, x_2, x_3))$ is also primitive recursive, since it is obtained by a substitution.

PROPOSITION 3.16

The following functions are primitive recursive.

(a) $x+y$
(b) $x \cdot y$
(c) x^y
(d) $\delta(x) = \begin{cases} x-1 & \text{if } x > 0 \\ 0 & \text{if } x = 0 \end{cases}$

δ is called the *predecessor* function.

(e) $x \dot{-} y = \begin{cases} x-y & \text{if } x \geqslant y \\ 0 & \text{if } x < y \end{cases}$

(f) $|x-y| = \begin{cases} x-y & \text{if } x \geqslant y \\ y-x & \text{if } x < y \end{cases}$

(g) $\mathrm{sg}(x) = \begin{cases} 0 & \text{if } x = 0 \\ 1 & \text{if } x \neq 0 \end{cases}$

(h) $\overline{\mathrm{sg}}(x) = \begin{cases} 1 & \text{if } x = 0 \\ 0 & \text{if } x \neq 0 \end{cases}$

(i) $x!$
(j) $\min(x,y) =$ minimum of x and y
(k) $\min(x_1, \ldots, x_n)$
(l) $\max(x,y) =$ maximum of x and y
(m) $\max(x_1, \ldots, x_n)$
(n) $\mathrm{rm}(x,y) =$ remainder upon division of y by x
(o) $\mathrm{qt}(x,y) =$ quotient upon division of y by x

Proof

(a) Recursion rule (V)

$$x+0 = x \qquad \text{or} \qquad f(x,0) = U_1^1(x)$$
$$x+(y+1) = N(x+y) \qquad f(x,y+1) = N(f(x,y))$$

(b)
$$x \cdot 0 = 0 \qquad \text{or} \qquad g(x,0) = Z(x)$$
$$x \cdot (y+1) = (x \cdot y) + x \qquad (x,y+1) = f(g(x,y),x)$$
where f is the addition function

(c)
$$x^0 = 1$$
$$x^{y+1} = (x^y) \cdot x$$

(d)
$$\delta(0) = 0$$
$$\delta(y+1) = y$$

(e)
$$x \dot{-} 0 = x$$
$$x \dot{-} (y+1) = \delta(x \dot{-} y)$$

(f) $|x-y| = (x \dot{-} y) + (y \dot{-} x)$ (substitution)

(g) $\mathrm{sg}(x) = x \dot{-} \delta(x)$ (substitution)

(h) $\overline{\text{sg}}(x) = 1 \dot{-} \text{sg}(x)$ (substitution)

(i) $0! = 1$

$(y+1)! = (y!) \cdot (y+1)$

(j) $\min(x,y) = x \dot{-} (x \dot{-} y)$

(k) Assume $\min(x_1, \ldots, x_n)$ already shown primitive recursive.

$$\min(x_1, \ldots, x_n, x_{n+1}) = \min(\min(x_1, \ldots, x_n), x_{n+1})$$

(l) $\max(x,y) = y + (x \dot{-} y)$

(m) $\max(x_1, \ldots, x_n, x_{n+1}) = \max(\max(x_1, \ldots, x_n), x_{n+1})$

(n) $\text{rm}(x,0) = 0$

$\text{rm}(x, y+1) = N(\text{rm}(x,y)) \cdot \text{sg}(|x - N(\text{rm}(x,y))|)$

(o) $\text{qt}(x,0) = 0$

$\text{qt}(x, y+1) = \text{qt}(x,y) + \overline{\text{sg}}(|x - N(\text{rm}(x,y))|)$

In justification of (n) and (o), note that, if q and r denote the quotient $\text{qt}(x,y)$ and remainder $\text{rm}(x,y)$ upon division of y by x, then $y = qx + r$ and $0 \leqslant r < x$. So, $y+1 = qx + (r+1)$. If $r+1 < x$ (that is, if $|x - N(\text{rm}(x,y))| > 0$), then the quotient $\text{qt}(x, y+1)$ and remainder $\text{rm}(x, y+1)$ upon division of $y+1$ by x are q and $r+1$, respectively. If $r+1 = x$ (that is, if $|x - N(\text{rm}(x,y))| = 0$), then $y+1 = (q+1)x$, and $\text{qt}(x, y+1)$ and $\text{rm}(x, y+1)$ are $q+1$ and 0, respectively.†

DEFINITIONS

$$\sum_{y<z} f(x_1, \ldots, x_n, y) = \begin{cases} 0 & \text{if } z = 0 \\ f(x_1, \ldots, x_n, 0) + \cdots + f(x_1, \ldots, x_n, z-1) & \text{if } z > 0 \end{cases}$$

$$\sum_{y \leqslant z} f(x_1, \ldots, x_n, y) = \sum_{y<z+1} f(x_1, \ldots, x_n, y)$$

$$\prod_{y<z} f(x_1, \ldots, x_n, y) = \begin{cases} 1 & \text{if } z = 0 \\ f(x_1, \ldots, x_n, 0) \cdot \ldots \cdot f(x_1, \ldots, x_n, z-1) & \text{if } z > 0 \end{cases}$$

$$\prod_{y \leqslant z} f(x_1, \ldots, x_n, y) = \prod_{y<z+1} f(x_1, \ldots, x_n, y)$$

These *bounded* sums and products are functions of $x_1, \ldots, x_n, z$. We can also define doubly bounded sums and products in terms of the ones already given; for example,

$$\sum_{u<y<v} f(x_1, \ldots, x_n, y) = f(x_1, \ldots, x_n, u+1) + \ldots + f(x_1, \ldots, x_n, v-1)$$

$$= \sum_{y < \delta(v \dot{-} u)} f(x_1, \ldots, x_n, y+u+1)$$

†Since one cannot divide by 0, the values of $\text{rm}(0,y)$ and $\text{qt}(0,y)$ have no intuitive significance. It can be easily shown by induction that the given definitions yield $\text{rm}(0,y) = y$ and $\text{qt}(0,y) = 0$.

PROPOSITION 3.17

If $f(x_1,\ldots,x_n,y)$ is primitive recursive (or recursive), then all the bounded sums and products defined above are also primitive recursive (or recursive).

Proof

Let $g(x_1,\ldots,x_n,z) = \sum_{y<z} f(x_1,\ldots,x_n,y)$. Then we have the following recursion:

$$g(x_1,\ldots,x_n,0) = 0$$
$$g(x_1,\ldots,x_n,z+1) = g(x_1,\ldots,x_n,z) + f(x_1,\ldots,x_n,z)$$

If $h(x_1,\ldots,x_n,z) = \sum_{y\leqslant z} f(x_1,\ldots,x_n,y)$, then

$$h(x_1,\ldots,x_n,z) = g(x_1,\ldots,x_n,z+1) \qquad \text{(substitution)}$$

The proofs for bounded products and doubly bounded sums and products are left as exercises.

Example

Let $\tau(x)$ be the number of divisors of x, if $x > 0$, and let $\tau(0) = 1$. (Thus, $\tau(x)$ is the number of divisors of x that are less than or equal to x.) Then τ is primitive recursive, since

$$\tau(x) = \sum_{y\leqslant x} \overline{\text{sg}}(\text{rm}(y,x))$$

Given expressions for number-theoretic relations, we can apply the connectives of the propositional calculus to them to obtain new expressions for relations. For example, if $R_1(x,y)$ and $R_2(x,u,v)$ are relations, then $R_1(x,y) \wedge R_2(x,u,v)$ is a new relation that holds for x,y,u,v when and only when both $R_1(x,y)$ and $R_2(x,u,v)$ hold. We shall use $(\forall y)_{y<z}R(x_1,\ldots,x_n,y)$ to express the relation: for all y, if y is less than z, then $R(x_1,\ldots,x_n,y)$ holds. We shall use $(\forall y)_{y\leqslant z}$, $(\exists y)_{y<z}$ and $(\exists y)_{y\leqslant z}$ in an analogous way; for example, $(\exists y)_{y<z}R(x_1,\ldots,x_n,y)$ means that there is some $y < z$ such that $R(x_1,\ldots,x_n,y)$ holds. We shall call $(\forall y)_{y<z}$, $(\forall y)_{y\leqslant z}(\exists y)_{y<z}$ and $(\exists y)_{y\leqslant z}$ *bounded quantifiers*. In addition, we define a *bounded* μ*-operator*:

$$\mu y_{y<z}R(x_1,\ldots,x_n,y) = \begin{cases} \text{the least } y < z \text{ for which } R(x_1,\ldots,x_n,y) \\ \text{holds if there is such a } y \\ z \text{ otherwise} \end{cases}$$

The value z is chosen in the second case because it is more convenient in later proofs; this choice has no intuitive significance. We also define $\mu y_{y\leqslant z}R(x_1,\ldots,x_n,y)$ to be $\mu y_{y<z+1}R(x_1,\ldots,x_n,y)$.

A relation $R(x_1,\ldots,x_n)$ is said to be *primitive recursive* (or recursive) if and only if its characteristic function $C_R(x_1,\ldots,x_n)$ is primitive recursive (or re-

cursive). In particular, a set A of natural numbers is primitive recursive (or recursive) if and only if its characteristic function $C_A(x)$ is primitive recursive (or recursive).

Examples

1. The relation $x_1 = x_2$ is primitive recursive. Its characteristic function is $\mathrm{sg}(|x_1 - x_2|)$, which is primitive recursive, by Proposition 3.16(f,g).
2. The relation $x_1 < x_2$ is primitive recursive, since its characteristic function is $\overline{\mathrm{sg}}(x_2 \dot{-} x_1)$, which is primitive recursive, by Proposition 3.16(e,h).
3. The relation $x_1 | x_2$ is primitive recursive, since its characteristic function is $\mathrm{sg}(\mathrm{rm}(x_1, x_2))$.
4. The relation $\mathrm{Pr}(x), (x)$ is a prime, is primitive recursive, since $C_{\mathrm{Pr}}(x) = \mathrm{sg}(|\tau(x) - 2|)$. Note that an integer is a prime if and only if it has exactly two divisors; recall that $\tau(0) = 1$.

PROPOSITION 3.18

Relations obtained from primitive recursive (or recursive) relations by means of the propositional connectives and the bounded quantifiers are also primitive recursive (or recursive). Also, applications of the bounded μ-operators $\mu y_{y<z}$ and $\mu y_{y\leqslant z}$ lead from primitive recursive (or recursive) relations to primitive recursive (or recursive) functions.

Proof

Assume $R_1(x_1, \ldots, x_n)$ and $R_2(x_1, \ldots, x_n)$ are primitive recursive (or recursive) relations. Then the characteristic functions C_{R_1} and C_{R_2} are primitive recursive (or recursive). But $C_{\neg R_1}(x_1, \ldots, x_n) = 1 \dot{-} C_{R_1}(x_1, \ldots, x_n)$; hence $\neg R_1$ is primitive recursive (or recursive). Also, $C_{R_1 \vee R_2}(x_1, \ldots, x_n) = C_{R_1}(x_1, \ldots, x_n) \cdot C_{R_2}(x_1, \ldots, x_n)$; so, $R_1 \vee R_2$ is primitive recursive (or recursive). Since all propositional connectives are definable in terms of $\neg$ and $\vee$, this takes care of them. Now, assume $R(x_1, \ldots, x_n, y)$ is primitive recursive (or recursive). If $Q(x_1, \ldots, x_n, z)$ is the relation $(\exists y)_{y<z} R(x_1, \ldots, x_n, y)$, then it is easy to verify that $C_Q(x_1, \ldots, x_n, z) = \Pi_{y<z} C_R(x_1, \ldots, x_n, y)$, which, by Proposition 3.17, is primitive recursive (or recursive). The bounded quantifier $(\exists y)_{y\leqslant z}$ is equivalent to $(\exists y)_{y<z+1}$, which is obtainable from $(\exists y)_{y<z}$ by substitution. Also, $(\forall y)_{y<z}$ is equivalent to $\neg(\exists y)_{y<z}\neg$, and $(\forall y)_{y\leqslant z}$ is equivalent to $\neg(\exists y)_{y\leqslant z}\neg$. Doubly bounded quantifiers, such as $(\exists y)_{u<y<v}$, can be defined by substitution, using the bounded quantifiers already mentioned. Finally, $\Pi_{u\leqslant y} C_R(x_1, \ldots, x_n, u)$ has the value 1 for all y such that $R(x_1, \ldots, x_n, u)$ is false for all $u \leqslant y$; it has the value 0 as soon as there is some $u \leqslant y$ such that $R(x_1, \ldots, x_n, u)$ holds. Hence, $\sum_{y<z}(\Pi_{u\leqslant y} C_R(x_1, \ldots, x_n, u))$

counts the number of integers from 0 up to but not including the first $y < z$ such that $R(x_1, \ldots, x_n, y)$ holds and is z if there is no such y; thus, it is equal to $\mu y_{y<z} R(x_1, \ldots, x_n, y)$ and so the latter function is primitive recursive (or recursive) by Proposition 3.17.

Examples

1. Let $p(x)$ be the x_{th} prime number in ascending order. Thus, $p(0) = 2$, $p(1) = 3$, $p(2) = 5$, and so on. We shall write p_x instead of $p(x)$. Then p_x is a primitive recursive function. In fact,

$$p_0 = 2$$
$$p_{x+1} = \mu y_{y \leqslant (p_x)!+1}(p_x < y \wedge \Pr(y))$$

   Notice that the relation $u < y \wedge \Pr(y)$ is primitive recursive. Hence, by Proposition 3.18, the function $\mu y_{y \leqslant v}(u < y \wedge \Pr(y))$ is a primitive recursive function $g(u, v)$. If we substitute the primitive recursive functions z and $z! + 1$ for u and v, respectively, in $g(u, v)$, we obtain the primitive recursive function

$$h(z) = \mu y_{y \leqslant z!+1}(z < y \wedge \Pr(y))$$

   and the right-hand side of the second equation above is $h(p_x)$; hence, we have an application of the recursion rule (V). The bound $(p_x)! + 1$ on the first prime after p_x is obtained from Euclid's proof of the infinitude of primes (see Exercise 3.23).

2. Every positive integer x has a unique factorization into prime powers: $x = p_0^{a_0} p_1^{a_1} \ldots p_k^{a_k}$. Let us denote by $(x)_j$ the exponent a_j in this factorization. If $x = 1, (x)_j = 1$ for all j. If $x = 0$, we arbitrarily let $(x)_j = 0$ for all j. Then the function $(x)_j$ is primitive recursive, since $(x)_j = \mu y_{y<x}(p_j^y | x \wedge \neg(p_j^{y+1} | x))$.

3. For $x > 0$, let $\ell h(x)$ be the number of non-zero exponents in the factorization of x into powers of primes, or, equivalently, the number of distinct primes that divide x. Let $\ell h(0) = 0$. Then ℓh is primitive recursive. To see this, let $R(x, y)$ be the primitive recursive relation $\Pr(y) \wedge y | x \wedge x \neq 0$. Then $\ell h(x) = \sum_{y \leqslant x} \overline{\mathrm{sg}}(C_R(x, y))$. Note that this yields the special cases $\ell h(0) = \ell h(1) = 0$. The expression '$\ell h(x)$' should be read 'length of x'.

4. If the number $x = 2^{a_0} 3^{a_1} \ldots p_k^{a_k}$ is used to 'represent' or 'encode' the sequence of positive integers $a_0, a_1, \ldots, a_k$, and $y = 2^{b_0} 3^{b_1} \ldots p_m^{b_m}$ 'represents' the sequence of positive integers $b_0, b_1, \ldots, b_m$, then the number

$$x * y = 2^{a_0} 3^{a_1} \ldots p_k^{a_k} p_{k+1}^{b_0} p_{k+2}^{b_1} \cdots p_{k+1+m}^{b_m}$$

   'represents' the new sequence $a_0, a_1, \ldots, a_k, b_0, b_1, \ldots, b_m$ obtained by juxtaposing the two sequences. Note that $\ell h(x) = k + 1$, which is the

length of the first sequence, $\ell h(y) = m + 1$, which is the length of the second sequence, and $b_j = (y)_j$. Hence,

$$x * y = x \cdot \prod_{j<\ell h(y)} (p_{\ell h(x)+j})^{(y)_j}$$

and, thus, $*$ is a primitive recursive function, called the *juxtaposition* function. It is not difficult to show that $x * (y * z) = (x * y) * z$ as long as $y \neq 0$ (which will be the only case of interest to us). Therefore, there is no harm in omitting parentheses when writing two or more applications of $*$. Also observe that $x * 0 = x * 1 = x$.

Exercises

3.16 Assume that $R(x_1, \ldots, x_n, y)$ is a primitive recursive (or recursive) relation. Prove the following:

(a) $(\exists y)_{u<y<v} R(x_1, \ldots, x_n, y)$, $(\exists y)_{u \leqslant y \leqslant v} R(x_1, \ldots, x_n, y)$ and $(\exists y)_{u \leqslant y<v} R(x_1, \ldots, x_n, y)$ are primitive (or recursive) relations.

(b) $\mu y_{u<y<v} R(x_1, \ldots, x_n, y)$, $\mu y_{u \leqslant y \leqslant v} R(x_1, \ldots, x_n, y)$ and $\mu y_{u \leqslant y<v} R(x_1, \ldots, x_n, y)$ are primitive recursive (or recursive) functions.

(c) If, for all natural numbers $x_1, \ldots, x_n$, there exists a natural number y such that $R(x_1, \ldots, x_n, y)$, then the function $f(x_1, \ldots, x_n) = \mu y R(x_1, \ldots, x_n, y)$ is recursive. [*Hint*: Apply the restricted μ-operator to $C_R(x_1, \ldots, x_n, y)$.]

3.17

(a) Show that the intersection, union and complement of primitive recursive (or recursive) sets are also primitive recursive (or recursive).

(b) Show that every finite set is primitive recursive.

3.18 Prove that a function $f(x_1, \ldots, x_n)$ is recursive if and only if its representing relation $f(x_1, \ldots, x_n) = y$ is a recursive relation.

3.19 Let $[\sqrt{n}]$ denote the greatest integer less than or equal to $\sqrt{n}$, and let $\Pi(n)$ denote the number of primes less than or equal to n. Show that $[\sqrt{n}]$ and $\Pi(n)$ are primitive recursive.

3.20 Let e be the base of the natural logarithms. Show that $[ne]$, the greatest integer less than or equal to ne, is a primitive recursive function.

3.21 Let $\mathrm{RP}(y, z)$ hold if and only if y and z are relatively prime, that is, y and z have no common factor greater than 1. Let $\varphi(n)$ be the number of positive integers less than or equal to n that are relatively prime to n. Prove that RP and φ are primitive recursive.

3.22 Show that, in the definition of the primitive recursive functions, one need not assume that $Z(x) = 0$ is one of the initial functions.

3.23 Prove that $p_{k+1} \leqslant (p_0 p_1 \ldots p_k) + 1$. Conclude that $p_{k+1} \leqslant p_k! + 1$.

For use in the further study of recursive functions, we prove the following theorem on definition by cases.

PROPOSITION 3.19

Let

$$f(x_1,\ldots,x_n)=\begin{cases} g_1(x_1,\ldots,x_n) & \text{if } R_1(x_1,\ldots,x_n) \text{ holds} \\ g_2(x_1,\ldots,x_n) & \text{if } R_2(x_1,\ldots,x_n) \text{ holds} \\ \vdots & \\ g_k(x_1,\ldots,x_n) & \text{if } R_k(x_1,\ldots,x_n) \text{ holds} \end{cases}$$

If the functions $g_1,\ldots,g_k$ and the relations $R_1,\ldots,R_k$ are primitive recursive (or recursive), and if, for any $x_1,\ldots,x_n$, exactly one of the relations $R_1(x_1,\ldots,x_n),\ldots,R_k(x_1,\ldots,x_n)$ is true, then f is primitive recursive (or recursive).

Proof

$$f(x_1,\ldots,x_n)=g_1(x_1,\ldots,x_n)\cdot\overline{\mathrm{sg}}(C_{R_1}(x_1,\ldots,x_n))+\ldots+ g_k(x_1,\ldots,x_n)\cdot\overline{\mathrm{sg}}(C_{R_k}(x_1,\ldots,x_n)).$$

Exercises

3.24 Show that in Proposition 3.19 it is not necessary to assume that R_k is primitive recursive (or recursive).

3.25 Let

$$f(x)=\begin{cases} x^2 & \text{if } x \text{ is even} \\ x+1 & \text{if } x \text{ is odd} \end{cases}$$

Prove that f is primitive recursive.

3.26 Let

$$h(x)=\begin{cases} 2 & \text{if Goldbach's conjecture is true} \\ 1 & \text{if Goldbach's conjecture is false} \end{cases}$$

Is h primitive recursive?

It is often important to have available a primitive recursive one–one correspondence between the set of ordered pairs of natural numbers and the set of natural numbers. We shall enumerate the pairs as follows:

$$(0,0),\quad (0,1),(1,0),(1,1),\quad (0,2),(2,0),(1,2),(2,1),(2,2),\quad \ldots$$

After we have enumerated all the pairs having components less than or equal to k, we then add a new group of all the new pairs having components less than or equal to $k+1$ in the following order: $(0,k+1),(k+1,0),(1,k+1),(k+1,1),\ldots,(k,k+1),(k+1,k),(k+1,k+1)$. If $x<y$, then (x,y) occurs before (y,x) and both are in the $(y+1)$th group. (Note that we start from 1 in counting groups.) The first y groups contain y^2 pairs, and (x,y) is the $(2x+1)$th pair in the $(y+1)$th group. Hence, (x,y) is the

$(y^2 + 2x + 1)$th pair in the ordering, and (y,x) is the $(y^2 + 2x + 2)$th pair. On the other hand, if $x = y$, (x,y) is the $((x+1)^2)$th pair. This justifies the following definition, in which $\sigma^2(x,y)$ denotes the place of the pair (x,y) in the above enumeration, with $(0,0)$ considered to be in the 0th place:

$$\sigma^2(x,y) = \mathrm{sg}(x \dot{-} y) \cdot (x^2 + 2y + 1) + \overline{\mathrm{sg}}(x \dot{-} y) \cdot (y^2 + 2x)$$

Clearly, σ^2 is primitive recursive.

Let us define inverse functions σ_1^2 and σ_2^2 such that $\sigma_1^2(\sigma^2(x,y)) = x, \sigma_2^2(\sigma^2(x,y)) = y$ and $\sigma^2(\sigma_1^2(z), \sigma_2^2(z)) = z$. Thus, $\sigma_1^2(z)$ and $\sigma_2^2(z)$ are the first and second components of the zth ordered pair in the given enumeration. Note first that $\sigma_1^2(0) = 0, \sigma_2^2(0) = 0$,

$$\sigma_1^2(n) = \begin{cases} \sigma_2^2(n) & \text{if } \sigma_1^2(n) < \sigma_2^2(n) \\ \sigma_2^2(n) + 1 & \text{if } \sigma_1^2(n) > \sigma_2^2(n) \\ 0 & \text{if } \sigma_1^2(n) = \sigma_2^2(n) \end{cases}$$

and

$$\sigma_2^2(n+1) = \begin{cases} \sigma_1^2(n) & \text{if } \sigma_1^2(n) \neq \sigma_2^2(n) \\ \sigma_1^2(n) + 1 & \text{if } \sigma_1^2(n) = \sigma_2^2(n) \end{cases}$$

Hence,

$$\begin{aligned} \sigma_1^2(n+1) &= \sigma_2^2(n) \cdot (\mathrm{sg}(\sigma_2^2(n) \dot{-} \sigma_1^2(n))) + (\sigma_2^2(n) + 1) \cdot (\mathrm{sg}(\sigma_1^2(n) \dot{-} \sigma_2^2(n))) \\ &= \varphi(\sigma_1^2(n), \sigma_2^2(n)) \\ \sigma_2^2(n+1) &= \sigma_1^2(n) \cdot (\mathrm{sg}(|\sigma_2^2(n) - \sigma_1^2(n)|)) + (\sigma_1^2(n) + 1) \cdot (\overline{\mathrm{sg}}(|\sigma_1^2(n) - \sigma_2^2(n)|)) \\ &= \psi(\sigma_1^2(n), \sigma_2^2(n)) \end{aligned}$$

where φ and ψ are primitive recursive functions. Thus, σ_1^2 and σ_2^2 are defined recursively at the same time. We can show that σ_1^2 and σ_2^2 are primitive recursive in the following devious way. Let $h(u) = 2^{\sigma_1^2(u)} 3^{\sigma_2^2(u)}$. Now, h is primitive recursive, since $h(0) = 2^{\sigma_1^2(0)} 3^{\sigma_2^2(0)} = 2^0 \cdot 3^0 = 1$, and $h(n+1) = 2^{\sigma_1^2(n+1)} 3^{\sigma_2^2(n+1)} = 2^{\varphi(\sigma_1^2(n), \sigma_2^2(n))} 3^{\psi(\sigma_1^2(n), \sigma_2^2(n))} = 2^{\varphi((h(n))_0, (h(n))_1)} \; 3^{\psi((h(n))_0, (h(n))_1)}$. Remembering that the function $(x)_i$ is primitive recursive (see Example 2 on page 181), we conclude by recursion rule (V) that h is primitive recursive. But $\sigma_1^2(x) = (h(x))_0$ and $\sigma_2^2(x) = (h(x))_1$. By substitution, σ_1^2 and σ_2^2 are primitive recursive.

One–one primitive recursive correspondences between all n-tuples of natural numbers and all natural numbers can be defined step – by – step, using induction on n. For $n = 2$, it has already been done. Assume that, for $n = k$, we have primitive recursive functions $\sigma^k(x_1, \ldots, x_k), \sigma_1^k(x), \ldots, \sigma_k^k(x)$ such that $\sigma_i^k(\sigma^k(x_1, \ldots, x_k)) = x_i$ for $1 \leqslant i \leqslant k$, and $\sigma^k(\sigma_1^k(x), \ldots, \sigma_k^k(x) = x$. Now, for $n = k + 1$, define $\sigma^{k+1}(x_1, \ldots, x_k, x_{k+1}) = \sigma^2(\sigma^k(x_1, \ldots, x_k), x_{k+1})$, $\sigma_i^{k+1}(x) = \sigma_i^k(\sigma_1^2(x))$ for $1 \leqslant i \leqslant k$ and $\sigma_{k+1}^{k+1}(x) = \sigma_2^2(x)$. Then $\sigma^{k+1}, \sigma_1^{k+1}, \ldots, \sigma_{k+1}^{k+1}$ are all primitive recursive, and we leave it as an exercise to verify that $\sigma_i^{k+1}(\sigma^{k+1}(x_1, \ldots, x_{k+1})) = x_i$ for $1 \leqslant i \leqslant k+1$, and $\sigma^k(\sigma_1^{k+1}(x), \ldots, \sigma_{k+1}^{k+1}(x)) = x$.

It will be essential in later work to define functions by a recursion in which the value of $f(x_1, \cdot, x_n, y+1)$ depends not only upon $f(x_1, \ldots, x_n, y)$ but also upon several or all values of $f(x_1, \ldots, x_n, u)$ with $u \leqslant y$. This type of recursion is called a *course-of-values* recursion. Let $f\#(x_1, \ldots, x_n, y) = \prod_{u<y} p_u^{f(x_1,\ldots,x_n,u)}$. Note that f can be obtained from $f\#$ as follows: $f(x_1, \ldots, x_n, y) = (f\#(x_1, \ldots, x_n, y+1))_y$.

PROPOSITION 3.20 (COURSE-OF-VALUES RECURSION)

If $h(x_1, \ldots, x_n, y, z)$ is primitive recursive (or recursive) and $f(x_1, \ldots, x_n, y) = h(x_1, \ldots, x_n, y, f\#(x_1, \ldots, x_n, y))$, then f is primitive recursive (or recursive).

Proof

$$f\#(x_1, \ldots, x_n, 0) = 1$$
$$f\#(x_1, \ldots, x_n, y+1) = f\#(x_1, \ldots, x_n, y) \cdot p_y^{f(x_1,\ldots,x_n,y)}$$
$$= f\#(x_1, \ldots, x_n, y) \cdot p_y^{h(x_1,\ldots,x_n,y,f\#(x_1,\ldots,x_n,y))}$$

Thus, by the recursion rule, $f\#$ is primitive recursive (or recursive), and $f(x_1, \ldots, x_n, y) = (f\#(x_1, \ldots, x_n, y+1))_y$.

Example
The Fibonacci sequence is defined as follows: $f(0) = 1, f(1) = 1$, and $f(k+2) = f(k) + f(k+1)$ for $k \geqslant 0$. Then f is primitive recursive, since

$$f(n) = \overline{\text{sg}}(n) + \overline{\text{sg}}(|n-1|) + ((f\#(n))_{n\dot{-}1} + (f\#(n))_{n\dot{-}2}) \cdot \text{sg}(n\dot{-}1)$$

The function

$$h(y, z) = \overline{\text{sg}}(y) + \overline{\text{sg}}(|y-1|) + ((z)_{y\dot{-}1} + (z)_{y\dot{-}2}) \cdot \text{sg}(y\dot{-}1)$$

is primitive recursive, and $f(n) = h(n, f\#(n))$.

Exercise

3.27 Let $g(0) = 2, g(1) = 4$, and $g(k+2) = 3g(k+1) - (2g(k) + 1)$. Show that g is primitive recursive.

COROLLARY 3.21 (COURSE-OF-VALUES RECURSION FOR RELATIONS)

If $H(x_1, \ldots, x_n, y, z)$ is a primitive recursive (or recursive) relation and $R(x_1, \ldots, x_n, y)$ holds if and only if $H(x_1, \cdots, x_n, y, (C_R)\#(x_1, \cdots, x_n, y))$,

where C_R is the characteristic function of R, then R is primitive recursive (or recursive).

Proof

$C_R(x_1,\ldots,x_n,y) = C_H(x_1,\ldots,x_n,y,(C_R)\#(x_1,\ldots,x_n,y))$. Since C_H is primitive recursive (or recursive), Proposition 3.20 implies that C_R is primitive recursive (or recursive) and, therefore, so is R.

Proposition 3.20 and Corollary 3.21 will be drawn upon heavily in what follows. They are applicable whenever the value of a function or relation for y is defined in terms of values for arguments less than y (by means of a primitive recursive or recursive function or relation). Notice in this connection that $R(x_1,\ldots,x_n,u)$ is equivalent to $C_R(x_1,\ldots,x_n,u) = 0$, which, in turn, for $u < y$, is equivalent to $((C_R)\#(x_1,\ldots,x_n,y))_u = 0$.

Exercises

3.28 Prove that the set of recursive functions is denumerable.

3.29 If $f_0,f_1,f_2,\ldots$ is an enumeration of all primitive recursive functions (or all recursive functions) of one variable , prove that the function $f_x(y)$ is not primitive recursive (or recursive).

LEMMA 3.22 (GÖDEL'S β-FUNCTION)

Let $\beta(x_1,x_2,x_3) = \mathrm{rm}(1 + (x_3 + 1)\cdot x_2, x_1)$. Then β is primitive recursive, by Proposition 3.16(n). Also , β is strongly representable in S by the following wf $\mathrm{Bt}(x_1,x_2,x_3,y)$:

$$(\exists \mathrm{W})(x_1 = (1 + (x_3 + 1)\cdot x_2)\cdot w + y \wedge y < 1 + (x_3 + 1)\cdot x_2)$$

Proof

By Proposition 3.11 $\vdash_{\mathrm{S}} (\exists_1 y)\mathrm{Bt}(x_1,x_2,x_3,y)$. Assume $\beta(k_1,k_2,k_3) = m$. Then $k_1 = (1 + (k_3 + 1)\cdot k_2)\cdot k + m$ for some k, and $m < 1 + (k_3 + 1)\cdot k_2$. So, $\vdash_{\mathrm{S}} \overline{k}_1 = (\overline{1} + (\overline{k}_3 + \overline{1})\cdot \overline{k}_2)\cdot \overline{k} + \overline{m}$, by Proposition 3.6(a). Moreover, $\vdash_{\mathrm{S}} \overline{m} < \overline{1} + (\overline{k}_3 + \overline{1})\cdot \overline{k}_2$ by the expressibility of $<$ and Proposition 3.6(a). Hence, $\vdash_{\mathrm{S}} \overline{k}_1 = (\overline{1} + (\overline{k}_3 + \overline{1})\cdot \overline{k}_2)\cdot \overline{k} + \overline{m} \wedge \overline{m} < \overline{1} + (\overline{k}_3 + \overline{1})\cdot \overline{k}_2$ from which by rule $E4$, $\vdash_{\mathrm{S}} \mathrm{Bt}(\overline{k}_1,\overline{k}_2,\overline{k}_3,\overline{m})$. Thus, Bt strongly represents β in S.

LEMMA 3.23

For any sequence of natural numbers $k_0,k_1,\ldots,k_n$, there exist natural numbers b and c such that $\beta(b,c,i) = k_i$ for $0 \leqslant i \leqslant n$.

Proof

Let $j = \max(n, k_0, k_1, \ldots, k_n)$ and let $c = j!$. Consider the numbers $u_i = 1 + (i+1)c$ for $0 \leqslant i \leqslant n$; no two of them have a factor in common other than 1. In fact, if p were a prime dividing both $1 + (i+1)c$ and $1 + (m+1)c$ with $0 \leqslant i < m \leqslant n$, then p would divide their difference $(m-i)c$. Now, p does not divide c, since, in that case p would divide both $(i+1)c$ and $1 + (i+1)c$, and so would divide 1, which is impossible. Hence, p also does not divide $(m-i)$; for $m - i \leqslant n \leqslant j$ and so, $m - i$ divides $j! = c$. If p divided $m - i$, then p would divide c. Therefore, p does not divide $(m-i)c$, which yields a contradiction. Thus, the numbers $u_i, 0 \leqslant i \leqslant n$, are relatively prime in pairs. Also, for $0 \leqslant i \leqslant n, k_i \leqslant j \leqslant j! = c < 1 + (i+1)c = u_i$; that is, $k_i < u_i$. Now, by the Chinese remainder theorem (see Exercise 3.30), there is a number $b < u_0 u_1 \ldots u_n$ such that $\mathrm{rm}(u_i, b) = k_i$ for $0 \leqslant i \leqslant n$. But $\beta(b, c, i) = \mathrm{rm}(1 + (i+1)c, b) = \mathrm{rm}(u_i, b) = k_i$.

Lemmas 3.22 and 3.23 enable us to express within S assertions about finite sequences of natural numbers, and this ability is crucial in part of the proof of the following fundamental theorem.

PROPOSITION 3.24

Every recursive function is representable in S.

Proof

The initial functions Z, N and U_i^n are representable in S, by Examples 1–3 on page 172. The substitution rule (IV) does not lead out of the class of representable functions, by Example 4 on page 172.

For the recursion rule (V), assume that $g(x_1, \ldots, x_n)$ and $h(x_1, \ldots, x_n, y, z)$ are representable in S by wfs $\mathscr{B}(x_1, \ldots, x_{n+1})$ and $\mathscr{C}(x_1, \ldots, x_{n+3})$, respectively, and let

$$(I) \qquad f(x_1, \ldots, x_n, 0) = g(x_1, \ldots, x_n)$$
$$f(x_1, \ldots, x_n, y+1) = h(x_1, \ldots, x_n, y, f(x_1, \ldots, x_n, y))$$

Now, $f(x_1, \ldots, x_n, y) = z$ if and only if there is a finite sequence of numbers $b_0, \ldots, b_y$ such that $b_0 = g(x_1, \ldots, x_n)$, $b_{w+1} = h(x_1, \ldots, x_n, w, b_w)$ for $w + 1 \leqslant y$, and $b_y = z$. But, by Lemma 3.23, reference to finite sequences can be paraphrased in terms of the function β and, by Lemma 3.22, β is representable in S by the wf $\mathrm{Bt}(x_1, x_2, x_3, y)$.

We shall show that $f(x_1, \ldots, x_n, x_{n+1})$ is representable in S by the following wf $\mathscr{D}(x_1, \ldots, x_{n+2})$:

$$(\exists u)(\exists v)[((\exists w)(\mathrm{Bt}(u, v, 0, w) \wedge \mathscr{B}(x_1, \ldots, x_n, w))) \wedge \mathrm{Bt}(u, v, x_{n+1}, x_{n+2})$$
$$\wedge (\forall w)(w < x_{n+1} \Rightarrow (\exists y)(\exists z)(Bt(u, v, w, y) \wedge \mathrm{Bt}(u, v, w', z) \wedge \mathscr{C}(x_1, \ldots, x_n, w, y, z)))]$$

(i) First, assume that $f(x_1,\ldots,x_n,p)=m$. We wish to show that $\vdash_S \mathscr{D}(\overline{k}_1,\ldots,\overline{k}_n,\overline{p},\overline{m})$. If $p=0$, then $m=g(k_1,\ldots,k_n)$. Consider the sequence consisting of m alone. By Lemma 3.23, there exist b and c such that $\beta(b,c,0)=m$. Hence, by Lemma 3.22,

$$(\mathfrak{X}) \qquad \vdash_S \mathrm{Bt}(\overline{b},\overline{c},0,\overline{m})$$

Also, since $m=g(k_1,\ldots,k_n)$, we have $\vdash_S \mathscr{B}(\overline{k}_1,\ldots,\overline{k}_n,\overline{m})$. Hence, by rule E4,

$$(\mathfrak{X}\mathfrak{X}) \qquad \vdash_S (\exists w)(\mathrm{Bt}(\overline{b},\overline{c},0,w)\wedge\mathscr{B}(\overline{k}_1,\ldots,\overline{k}_n,w))$$

In addition, since $\vdash_S w\nless 0$, a tautology and Gen yield

$$(\mathfrak{X}\mathfrak{X}\mathfrak{X}) \quad (\forall w)(w<0\Rightarrow(\exists y)(\exists z)(\mathrm{Bt}(\overline{b},\overline{c},w,y)\wedge(\mathrm{Bt}(\overline{b},\overline{c},w',z)\wedge\mathscr{C}(\overline{k}_1,\ldots,\overline{k}_n,w,y,z)))$$

Applying rule E4 to the conjunction of $(\mathfrak{X})$, $(\mathfrak{X}\mathfrak{X})$ and $(\mathfrak{X}\mathfrak{X}\mathfrak{X})$, we obtain $\vdash_S \mathscr{D}(\overline{k}_1,\ldots,\overline{k}_n,0,\overline{m})$. Now, for $p>0$, $f(k_1,\ldots,k_n,p)$ is calculated from the equations (I) in $p+1$ steps. Let $r_i=f(k_1,\ldots,k_n,i)$. For the sequence of numbers $r_0,\ldots,r_p$, there are, by Lemma 3.23, numbers b and c such that $\beta(b,c,i)=r_i$ for $0\leqslant i\leqslant p$. Hence, by Lemma 3.22, $\vdash_S \mathrm{Bt}(\overline{b},\overline{c},\overline{i},\overline{r}_i)$. In particular, $\beta(b,c,0)=r_0=f(k_1,\ldots,k_n,0)=g(k_1,\ldots,k_n)$. Therefore, $\vdash_S \mathrm{Bt}(\overline{b},\overline{c},0,\overline{r}_0)\wedge\mathscr{B}(\overline{k}_1,\ldots,\overline{k}_n,\overline{r}_0)$, and, by rule E4, (i) $\vdash_S (\exists w)(\mathrm{Bt}(\overline{b},\overline{c},0,w)\wedge\mathscr{B}(\overline{k}_1,\ldots,\overline{k}_n,w))$. Since $r_p=f(k_1,\ldots,k_n,p)=m$, we have $\beta(b,c,p)=m$. Hence, (ii) $\vdash_S \mathrm{Bt}(\overline{b},\overline{c},\overline{p},\overline{m})$. For $0<i\leq p-1$, $\beta(b,c,i)=r_i=f(k_1,\cdots,k_n,i)$ and $\beta(b,c,i+1)=r_{i+1}=f(k_1,\cdots,k_n,i+1)=h(k_1,\cdots,k_n,i,f(k_1,\cdots,k_n,i))=h(k_1,\cdots,k_n,i,r_i)$. Therefore, $\vdash_S \mathrm{Bt}(\overline{b},\overline{c},\overline{i},\overline{r}_i)\wedge\mathrm{Bt}(\overline{b},\overline{c},\overline{i}',\overline{r}_{i+1})\wedge\mathscr{C}(\overline{k}_1,\cdots,\overline{k}_n,\overline{i},\overline{r}_i,\overline{r}_{i+1})$. By Rule E4, $\vdash_S (\exists y)(\exists z)(\mathrm{Bt}(\overline{b},\overline{c},\overline{i},y)\wedge\mathrm{Bt}(\overline{b},\overline{c},\overline{i}',\overline{z})\wedge\mathscr{C}(\overline{k}_1,\cdots,\overline{k}_n,\overline{i},y,z))$ So, by Proposition 3.8(b′), (iii) $\vdash_S (\forall w)(w<\overline{p}\Rightarrow(\exists y)(\exists z)(\mathrm{Bt}(\overline{b},\overline{c},w,y)\wedge\mathrm{Bt}(\overline{b},\overline{c},w',z)\wedge\mathscr{C}(\overline{k}_1,\ldots,\overline{k}_n,w,y,z)))$. Then, applying rule E4 twice to the conjunction of (i), (ii) and (iii), we obtain $\vdash_S \mathscr{D}(\overline{k}_1,\cdots\overline{k}_n,\overline{p},\overline{m})$. Thus, we have verified clause 1 of the definition of representability (see page 171).

(ii) We must show that $\vdash_S (\exists_1 x_{n+2})\mathscr{D}(\overline{k}_1,\ldots,\overline{k}_n,\overline{p},x_{n+2})$. The proof is by induction on p in the metalanguage. Notice that, by what we have proved above, it suffices to prove only uniqueness. The case of $p=0$ is left as an easy exercise. Assume $\vdash_S (\exists_1 x_{n+2})\mathscr{D}(\overline{k}_1,\ldots,\overline{k}_n,\overline{p},x_{n+2})$. Let $\alpha=g(k_1,\ldots,k_n)$, $\beta=f(k_1,\ldots,k_n,p)$, and $\gamma=f(k_1,\ldots,k_n,p+1)=h(k_1,\ldots,k_n,p,\beta)$. Then

(1) $\vdash_S \mathscr{C}(\overline{k}_1,\ldots,\overline{k}_n,\overline{p},\overline{\beta},\overline{\gamma})$

(2) $\vdash_S \mathscr{B}(\overline{k}_1,\ldots,\overline{k}_n,\overline{\alpha})$

(3) $\vdash_S \mathscr{D}(\overline{k}_1,\ldots,\overline{k}_n,\overline{p},\overline{\beta})$

(4) $\vdash_S \mathscr{D}(\overline{k}_1,\ldots,\overline{k}_n,\overline{p+1},\overline{\gamma})$

(5) $\vdash_S (\exists_1 x_{n+2})\mathscr{D}(\overline{k}_1,\ldots,\overline{k}_n,\overline{p},x_{n+2})$

Assume

$$(6)\ \mathscr{D}(\bar{k}_1,\ldots,\bar{k}_n,\overline{p+1},x_{n+2})$$

We must prove $x_{n+2}=\bar{\gamma}$. From (6), by rule C,

(a) $(\exists w)(\mathrm{Bt}(b,c,0,w)\wedge\mathscr{B}(\bar{k}_1,\ldots,\bar{k}_n,w))$

(b) $\mathrm{Bt}(b,c,\overline{p+1},x_{n+2})$

(c) $(\forall w)(w<\overline{p+1}$

$\Rightarrow(\exists y)(\exists z)(\mathrm{Bt}(b,c,w,y)\wedge\mathrm{Bt}(b,c,w',z)\wedge\mathscr{C}(\bar{k}_1,\ldots,\bar{k}_n,w,y,z)))$

From (c),

(d) $(\forall w)(w<\bar{p}\Rightarrow(\exists y)(\exists z)(\mathrm{Bt}(b,c,w,y)\wedge\mathrm{Bt}(b,c,w',z)\wedge\mathscr{C}(\bar{k}_1,\ldots,\bar{k}_n,w,y,z)))$

From (c) by rule A4 and rule C,

$$(e)\ \mathrm{Bt}(b,c,\bar{p},d)\wedge\mathrm{Bt}(b,c,\overline{p+1},e)\wedge\mathscr{C}(\bar{k}_1,\ldots,\bar{k}_n,\bar{p},d,e)$$

From (a), (d), and (e),

$$(f)\ \ \mathscr{D}(\bar{k}_1,\ldots,\bar{k}_n,\bar{p},d)$$

From (f), (5) and (3),

$$(g)\ d=\bar{\beta}$$

From (e) and (g),

$$(h)\ \mathscr{C}(\bar{k}_1,\ldots,\bar{k}_n,\bar{p},\bar{\beta},e)$$

Since β represents h, we obtain from (1) and (h),

$$(i)\ \bar{\gamma}=e$$

From (e) and (i),

$$(j)\ \mathrm{Bt}(b,c,\overline{p+1},\bar{\gamma})$$

From (b), (j), and Lemma 3.22,

$$(k)\ x_{n+2}=\bar{\gamma}$$

This completes the induction.

The μ-operator (VI). Let us assume, that, for any $x_1,\ldots,x_n$, there is some y such that $g(x_1,\ldots,x_n,y)=0$, and let us assume g is representable in S by a wf $\mathscr{E}(x_1,\ldots,x_{n+2})$. Let $f(x_1,\ldots,x_n)=\mu y(g(x_1,\ldots,x_n,y)=0)$. Then we shall show that f is representable in S by the wf $\mathscr{F}(x_1,\ldots,x_{n+1})$:

$$\mathscr{E}(x_1,\ldots,x_{n+1},0)\wedge(\forall y)(y<x_{n+1}\Rightarrow\neg\mathscr{E}(x_1,\ldots,x_n,y,0))$$

Assume $f(k_1,\ldots,k_n)=m$. Then $g(k_1,\ldots,k_n,m)=0$ and, for $k<m$, $g(k_1,\ldots,k_n,k)\neq 0$. So, $\vdash_S\mathscr{E}(\bar{k}_1,\ldots,\bar{k}_n,\bar{m},0)$ and, for $k<m$, $\vdash_S\neg\mathscr{E}(\bar{k}_1,\cdots,\bar{k}_n,\bar{k},0)$. By Proposition 3.8(b′), $\vdash_S(\forall y)(y<\bar{m}\Rightarrow\neg\mathscr{E}(\bar{k}_1,\ldots,\bar{k}_n,y,0))$. Hence, $\vdash_S\mathscr{F}(\bar{k}_1,\ldots,\bar{k}_n,\bar{m})$. We must also show: $\vdash_S(\exists_1 x_{n+1})\mathscr{F}(\bar{k}_1,\ldots,\bar{k}_n,x_{n+1})$. It suffices to prove the uniqueness. Assume $\mathscr{E}(\bar{k}_1,\ldots,\bar{k}_n,u,0)\wedge(\forall y)(y<u\Rightarrow\neg\mathscr{E}(\bar{k}_1,\ldots,\bar{k}_n,y,0))$. By Proposition

3.7(o'), $\vdash_S \overline{m} < u \vee \overline{m} = u \vee u < \overline{m}$. Since $\vdash_S \mathscr{E}(\overline{k}_1, \ldots, \overline{k}_n, \overline{m}, 0)$, we cannot have $\overline{m} < u$. Since $\vdash_S (\forall y)(y < \overline{m} \Rightarrow \neg\mathscr{E}(\overline{k}_1, \ldots, \overline{k}_n, y, 0))$, we cannot have $u < \overline{m}$. Hence, $u = \overline{m}$. This shows the uniqueness.

Thus, we have proved that all recursive functions are representable in S.

COROLLARY 3.25

Every recursive relation is expressible in S.

Proof

Let $R(x_1, \ldots, x_n)$ be a recursive relation. Then its chracteristic function C_R is recursive. By Proposition 3.24, C_R is representable in S and, therefore, by Proposition 3.13, R is expressible in S.

Exercises

3.30[A]

(a) Show that, if a and b are relatively prime natural numbers, then there is a natural number c such that $ac \equiv 1 (\mathrm{mod}\, b)$. (Two numbers a and b are said to be *relatively prime* if their greatest common divisor is 1. In general, $x \equiv y (\mathrm{mod}\, z)$ means that x and y leave the same remainder upon division by z or, equivalently, that $x - y$ is divisible by z. This exercise amounts to showing that there exist integers u and v such that $1 = au + bv$.)

(b) Prove the Chinese remainder theorem: if $x_1, \ldots, x_k$ are relatively prime in pairs and $y_1, \ldots, y_k$ are any natural numbers, there is a natural number z such that $z \equiv y_1 (\mathrm{mod}\, x_1), \ldots, z \equiv y_k (\mathrm{mod}\, x_k)$. Moreover, any two such zs differ by a multiple of $x_1 \ldots x_k$. [*Hint*: Let $x = x_1 \ldots x_k$ and let $x = w_1 x_1 = w_2 x_2 = \ldots = w_k x_k$. Then, for $1 \leqslant j \leqslant k$, w_j is relatively prime to x_j and so, by (a), there is some z_j such that $w_j z_j \equiv 1 (\mathrm{mod}\, x_j)$. Now let $z = w_1 z_1 y_1 + w_2 z_2 y_2 + \ldots + w_k z_k y_k$. Then $z \equiv w_j z_j y_j \equiv y_j (\mathrm{mod}\, x_j)$. In addition, the difference between any two such solutions is divisible by each of $x_1, \ldots, x_k$ and hence by $x_1 \ldots x_k$.]

3.31 Call a relation $R(x_1, \ldots, x_n)$ *arithmetical* if it is the interpretation of some wf $\mathscr{B}(x_1, \ldots, x_n)$ in the language $\mathscr{L}_A$ of arithmetic with respect to the standard model. Show that every recursive relation is arithmetical. [*Hint:* Use Corollary 3.25.]

3.4 ARITHMETIZATION. GÖDEL NUMBERS

For an arbitrary first-order theory K, we correlate with each symbol u of K an odd positive integer $g(u)$, called the *Gödel number* of u, in the following manner:

$$g(() = 3, g()) = 5, g(,) = 7, g(\neg) = 9, g(\Rightarrow) = 11, g(\forall) = 13,$$
$$g(x_k) = 13 + 8k \quad \text{for } k \geqslant 1$$
$$g(a_k) = 7 + 8k \quad \text{for } k \geqslant 1$$
$$g(f_k^n) = 1 + 8(2^n 3^k) \quad \text{for } k, n \geqslant 1$$
$$g(A_k^n) = 3 + 8(2^n 3^k) \quad \text{for } k, n \geqslant 1$$

Clearly, every Gödel number of a symbol is an odd positive integer. Moreover, when divided by 8, $g(u)$ leaves a remainder of 5 when u is a variable, a remainder of 7 when u is an individual constant, a remainder of 1 when u is a function letter, and a remainder of 3 when u is a predicate letter. Thus, different symbols have different Gödel numbers.

Examples

$g(x_2) = 29,\ g(a_4) = 39,\ g(f_1^2) = 97,\ g(A_2^1) = 147$

Given an expression $u_0 u_1 \ldots u_r$, where each u_j is a symbol of K, we define its Gödel number $g(u_0 u_1 \ldots u_r)$ by the equation

$$g(u_0 u_1 \ldots u_r) = 2^{g(u_0)} 3^{g(u_1)} \ldots p_r^{g(u_r)}$$

where p_j denotes the jth prime number and we assume that $p_0 = 2$. For example,

$$g(A_1^2(x_1, x_2)) = 2^{g(A_1^2)} 3^{g(()} 5^{g(x_1)} 7^{g(,)} 11^{g(x_2)} 13^{g())}$$
$$= 2^{99} 3^3 5^{21} 7^7 11^{29} 13^5$$

Observe that different expressions have different Gödel numbers, by virtue of the uniqueness of the factorization of integers into primes. In addition, expressions have different Gödel numbers from symbols, since the former have even Gödel numbers and the latter odd Gödel numbers. Notice also that a single symbol, considered as an expression, has a different Gödel number from its Gödel number as a symbol. For example, the symbol x_1 has Gödel number 21, whereas the expression that consists of only the symbol x_1 has Gödel number 2^{21}.

If $e_0, e_1, \ldots, e_r$ is any finite sequence of expressions of K, we can assign a Gödel number to this sequence by setting

$$g(e_0, e_1, \ldots, e_r) = 2^{g(e_0)} 3^{g(e_1)} \ldots p_r^{g(e_r)}$$

Different sequences of expressions have different Gödel numbers. Since a Gödel number of a sequence of expressions is even and the exponent of 2 in its prime power factorization is also even, it differs from Gödel numbers of symbols and expressions. Remember that a proof in K is a certain kind of finite sequence of expressions and, therefore, has a Gödel number.

Thus, g is a one–one function from the set of symbols of K, expressions of K, and finite sequences of expressions of K, into the set of positive

integers. The range of g is not the whole set of positive integers. For example, 10 is not a Gödel number.

Exercises

3.32 Determine the objects that have the following Gödel numbers.

(a) 1944 (b) 49 (c) 15 (d) 13 824 (e) $2^{5}13^{11}5^{9}$

3.33 Show that, if n is odd, $4n$ is not a Gödel number.

3.34 Find the Gödel numbers of the following expressions.

(a) $f_1^1(a_1)$ (b) $((\forall x_3)(\neg A_1^2(a_1, x_3)))$

This method of associating numbers with symbols, expressions and sequences of expressions was originally devised by Gödel (1931) in order to *arithmetize* metamathematics,† that is, to replace assertions about a formal system by equivalent number-theoretic statements and then to express these statements within the formal system itself. This idea turned out to be the key to many significant problems in mathematical logic.

The assignment of Gödel numbers given here is in no way unique. Other methods are found in Kleene (1952, chap. X) and in Smullyan (1961, chap. 1, § 6).

DEFINITION

A theory K is said to have a *primitive recursive vocabulary* (or a *recursive vocabulary*) if the following properties are primitive recursive (or recursive):

(a) IC(x): x is the Gödel number of an individual constant of K;
(b) FL(x): x is the Gödel number of a function letter of K;
(c) PL(x): x is the Gödel number of a predicate letter of K.

REMARK

Any theory K that has only a finite number of individual constants, function letters, and predicate letters has a primitive recursive vocabulary. For example, if the individual constants of K are $a_{j_1}, a_{j_2}, \ldots, a_{j_n}$, then IC($x$) if and only if $x = 7 + 8_{j_1} \vee x = 7 + 8_{j_2} \vee \ldots \vee x = 7 + 8_{j_n}$. In particular, any theory

†An *arithmetization* of a theory K is a one–one function g from the set of symbols of K, expressions of K and finite sequences of expressions of K into the set of positive integers. The following conditions are to be satisfied by the function g: (1) g is effectively computable; (2) there is an effective procedure that determines whether any given positive integer m is in the range of g and, if m is in the range of g, the procedure finds the object x such that $g(x) = m$.

K in the language $\mathscr{L}_A$ of arithmetic has a primitive recursive vocabulary. So, S has a primitive recursive vocabulary.

PROPOSITION 3.26

Let K be a theory with a primitive recursive (or recursive) vocabulary. Then the following relations and functions (1–16) are primitive recursive (or recursive). In each case, we give first the notation and intuitive definition for the relation or function, and then an equivalent formula from which its primitive recursiveness (or recursiveness) can be deduced.

(1) EVbl(x): x is the Gödel number of an expression consisting of a variable, $(\exists z)_{z<x}(1 \leqslant z \wedge x = 2^{13+8z})$. By Proposition 3.18, this is primitive recursive.

EIC(x): x is the Gödel number of an expression consisting of an individual constant, $(\exists y)_{y<x}(\mathrm{IC}(y) \wedge x = 2^y)$ (Proposition 3.18).
EFL(x): x is the Gödel number of an expression consisting of a function letter, $(\exists y)_{y<x}(\mathrm{FL}(y) \wedge x = 2^y)$ (Proposition 3.18).
EPL(x): x is the Gödel number of an expression consisting of a predicate letter, $(\exists y)_{y<x}(\mathrm{PL}(y) \wedge x = 2^y)$ (Proposition 3.18).

(2) $\mathrm{Arg_T}(x) = (\mathrm{qt}(8, x\dot{-}1))_0$: If x is the Gödel number of a function letter f_j^n, then $\mathrm{Arg_T}(x) = n$. $\mathrm{Arg_T}(x)$ is primitive recursive.

$\mathrm{Arg_P}(x) = (\mathrm{qt}(8, x\dot{-}3))_0$: If x is the Gödel number of a predicate letter A_j^n, then $\mathrm{Arg_P}(x) = n$. $\mathrm{Arg_P}(x)$ is primitive recursive.

(3) Gd(x): x is the Gödel number of an expression of K, $\mathrm{EVbl}(x) \vee \mathrm{EIC}(x) \vee \mathrm{EFL}(x) \vee \mathrm{EPL}(x) \vee x = 2^3 \vee x = 2^5 \vee x = 2^7 \vee x = 2^9 \vee x = 2^{11} \vee x = 2^{13} \vee (\exists u)_{u<x}(\exists v)_{v<x}(x = u * v \wedge \mathrm{Gd}(u) \wedge \mathrm{Gd}(v))$. Use Corollary 3.21. Here, $*$ is the juxtaposition function defined in Example 4 on page 181.

(4) MP(x, y, z): The expression with Gödel number z is a direct consequence of the expressions with Gödel numbers x and y by modus ponens, $y = 2^3 * x * 2^{11} * z * 2^5 \wedge \mathrm{Gd}(x) \wedge \mathrm{Gd}(z)$.

(5) Gen(x, y): The expression with Gödel number y comes from the expression with Gödel number x by the generalization rule:

$$(\exists v)_{v<y}(\mathrm{EVbl}(v) \wedge y = 2^3 * 2^3 * 2^{13} * v * 2^5 * x * 2^5 \wedge \mathrm{Gd}(x))$$

(6) Trm(x): x is the Gödel number of a term of K. This holds when and only when either x is the Gödel number of an expression consisting of a variable or an individual constant, or there is a function letter f_k^n and terms $t_1, \ldots, t_n$ such that x is the Gödel number of $f_k^n(t_1, \ldots, t_n)$. The latter holds if and only if there is a sequence of $n+1$ expressions

$$f_k^n(\ f_k^n(t_1,\ f_k^n(t_1, t_2,\ \ \cdots\ \ f_k^n(t_1, \ldots, t_{n-1},\ f_k^n(t_1, \ldots, t_{n-1}, t_n)$$

the last of which, $f_k^n(t_1, \ldots, t_n)$, has Gödel number x. This sequence can be represented by its Gödel number y. Clearly, $y < 2^x 3^x \ldots p_n^x = (2 \cdot 3 \cdot \ldots \cdot p_n)^x < (p_n!)^x < (p_x!)^x$. Note that $\ell h(y) = n + 1$ and also that $n = \mathrm{Arg_T}((x)_0)$, since $(x)_0$ is the Gödel number of f_k^n. Hence, $\mathrm{Trm}(x)$ is equivalent to the following relation:

$$\begin{aligned}&\mathrm{EVbl}(x) \vee \mathrm{EIC}(x) \vee (\exists y)_{y<(p_x!)^x}[x = (y)_{\ell h(y)\dot{-}1} \wedge\\ &\ell h(y) = \mathrm{Arg_T}((x)_0) + 1 \wedge \mathrm{FL}(((y)_0)_0) \wedge ((y)_0)_1 = 3 \wedge\\ &\ell h((y)_0) = 2 \wedge (\forall u)_{u<\ell h(y)\dot{-}2}(\exists v)_{v<x}((y)_{u+1} = (y)_u * v * 2^7 \wedge \mathrm{Trm}(v)) \wedge\\ &(\exists v)_{v<x}((y)_{\ell h(y)\dot{-}1} = (y)_{\ell h(y)\dot{-}2} * v * 2^5 \wedge \mathrm{Trm}(v))]\end{aligned}$$

Thus, $\mathrm{Trm}(x)$ is primitive recursive (or recursive) by Corollary 3.21, since the formula above involves $\mathrm{Trm}(v)$ for only $v < x$. In fact, if we replace both occurrences of $\mathrm{Trm}(v)$ in the formula by $(z)_v = 0$, then the new formula defines a primitive recursive (or recursive) relation $\mathrm{H}(x, z)$, and $\mathrm{Trm}(x) \Leftrightarrow \mathrm{H}(x, (C_{\mathrm{Trm}})^{\#}(x))$. Therefore, Corollary 3.21 is applicable.

(7) $\mathrm{Atfml}(x)$: x is the Gödel number of an atomic wf of K. This holds if and only if there are terms $t_1, \ldots, t_n$ and a predicate letter A_k^n such that x is the Gödel number of $A_k^n(t_1, \ldots, t_n)$. The latter holds if and only if there is a sequence of $n + 1$ expressions

$$A_k^n(\ A_k^n(t_1,\ A_k^n(t_1, t_2,\ \ \cdots\ \ A_k^n(t_1, \ldots, t_{n-1},\ A_k^n(t_1, \ldots, t_{n-1}, t_n)$$

the last of which, $A_k^n(t_1, \ldots, t_n)$, has Gödel number x. This sequence of expressions can be represented by its Gödel number y. Clearly, $y < (p_x!)^x$ (as in (6) above) and $n = \mathrm{Arg_P}((x)_0)$. Thus, $\mathrm{Atfml}(x)$ is equivalent to the following:

$$\begin{aligned}&(\exists y)_{y<(p_x!)^x}[x = (y)_{\ell h(y)\dot{-}1} \wedge \ell h(y) = \mathrm{Arg_P}((x)_0) + 1 \wedge\\ &\mathrm{PL}(((y)_0)_0) \wedge ((y)_0)_1 = 3 \wedge \ell h((y)_0) = 2 \wedge\\ &(\forall u)_{u<\ell h(y)\dot{-}2}(\exists v)_{v<x}((y)_{u+1} = (y)_u * v * 2^7 \wedge \mathrm{Trm}(v)) \wedge\\ &(\exists v)_{v<x}((y)_{\ell h(y)\dot{-}1} = (y)_{\ell h(y)\dot{-}2} * v * 2^5 \wedge \mathrm{Trm}(v))]\end{aligned}$$

Hence, by Proposition 3.18, $\mathrm{Atfml}(x)$ is primitive recursive (or recursive)

(8) $\mathrm{Fml}(y)$: y is the Gödel number of a formula of K:

$$\begin{aligned}&\mathrm{Atfml}(y) \vee (\exists z)_{z<y}[(\mathrm{Fml}(z) \wedge y = 2^3 * 2^9 * z * 2^5) \vee\\ &(\mathrm{Fml}((z)_0) \wedge \mathrm{Fml}((z)_1) \wedge y = 2^3 * (z)_0 * 2^{11} * (z)_1 * 2^5) \vee\\ &(\mathrm{Fml}((z)_0) \wedge \mathrm{EVbl}((z)_1) \wedge y = 2^3 * 2^3 * 2^{13} * (z)_1 * 2^5 * (z)_0 * 2^5)]\end{aligned}$$

It is easy to verify that Corollary 3.21 is applicable.

(9) $\mathrm{Subst}(x, y, u, v)$: x is the Gödel number of the result of substituting in the expression with Gödel number y the term with Gödel number u for all free occurrences of the variable with Gödel number v:

$$\begin{aligned}&\mathrm{Gd}(y) \wedge \mathrm{Trm}(u) \wedge \mathrm{EVbl}(2^v) \wedge [(y = 2^v \wedge x = u) \vee\\ &(\exists w)_{w<y}(y = 2^w \wedge y \neq 2^v \wedge x = y) \vee\\ &(\exists z)_{z<y}(\exists w)_{w<y}(\mathrm{Fml}(w) \wedge y = 2^3 * 2^{13} * 2^v * 2^5 * w * z \wedge\\ &(\exists \alpha)_{\alpha<x}(x = 2^3 * 2^{13} * 2^v * 2^5 * w * \alpha \wedge \mathrm{Subst}(\alpha, z, u, v))) \vee\\ &((\neg(\exists z)_{z<y}(\exists w)_{w<y}(\mathrm{Fml}(w) \wedge y = 2^3 * 2^{13} * 2^v * 2^5 * w * z)) \wedge\\ &(\exists \alpha)_{\alpha<x}(\exists \beta)_{\beta<x}(\exists z)_{z<y}(1 < z \wedge y = 2^{(y)_0} * z \wedge x = \alpha * \beta \wedge\\ &\mathrm{Subst}(\alpha, 2^{(y)_0}, u, v) \wedge \mathrm{Subst}(\beta, z, u, v)))]\end{aligned}$$

Corollary 3.21 is applicable. The reader should verify that this formula actually captures the intuitive content of Subst(x, y, u, v).

(10) Sub(y, u, v): the Gödel number of the result of substituting the term with Gödel number u for all free occurrences in the expression with Gödel number y of the variable with Gödel number v:

$$\mathrm{Sub}(y, u, v) = \mu x_{x<(p_{lh\,y}!)^{up}}\mathrm{Subst}(u, y, u, v)$$

Therefore, Sub is primitive recursive (or recursive) by Proposition 3.18. (When the conditions on u, v and y are not met, Sub(y, u, v) is defined, but its value is of no interest.)

(11) Fr(y, v): y is the Gödel number of a wf or term of K that contains free occurrences of the variable with Gödel number v:

$$(\mathrm{Fml}(y) \vee \mathrm{Trm}(y)) \wedge \mathrm{EVbl}(2^v) \wedge \neg\mathrm{Subst}(y, y, 2^{13+8v}, v)$$

(That is, substitution in the wf or term with Gödel number y of a certain variable different from the variable with Gödel number v for all free occurrences of the variable with Gödel number v yields a different expression.)

(12) Ff(u, v, w): u is the Gödel number of a term that is free for the variable with Gödel number v in the wf with Gödel number w:

$$\begin{aligned}&\mathrm{Trm}(u) \wedge \mathrm{EVbl}(2^v) \wedge \mathrm{Fml}(w) \wedge [\mathrm{Atfml}(w)\\ &\wedge (\exists y)_{y<w}(w = 2^3 * 2^9 * y * 2^5 \wedge \mathrm{Ff}(u, v, y))\\ &\vee (\exists y)_{y<w}(\exists z)_{z<w}(w = 2^3 * y * 2^{11} * z * 2^5\\ &\wedge \mathrm{Ff}(u, v, y) \wedge \mathrm{Ff}(u, v, z)) \vee\\ &(\exists y)_{y<w}(\exists z)_{z<w}(w = 2^3 * 2^3 * 2^{13} * 2^z * 2^5 * y * 2^5\\ &\wedge \mathrm{EVbl}(2^z) \wedge (z \neq v \Rightarrow \mathrm{Ff}(u, v, y)\\ &\wedge (\mathrm{Fr}(u, z) \Rightarrow \neg\mathrm{Fr}(y, v))))]\end{aligned}$$

Use Corollary 3.21 again.

(13) (a) $\mathrm{Ax}_1(x)$: x is the Gödel number of an instance of axiom schema (A1):

$$\begin{aligned}&(\exists u)_{u<x}(\exists v)_{v<x}(\mathrm{Fml}(u) \wedge \mathrm{Fml}(v)\\ &\wedge x = 2^3 * u * 2^{11} * 2^3 * v * 2^{11} * u * 2^5 * 2^5)\end{aligned}$$

(b) $\mathrm{Ax}_2(x)$: x is the Gödel number of an instance of axiom schema (A2):

$$(\exists u)_{u<x}(\exists v)_{v<x}(\exists w)_{w<x}(\mathrm{Fml}(u) \wedge \mathrm{Fml}(v) \wedge \mathrm{Fml}(w)$$
$$\wedge x = 2^3 * 2^3 * u * 2^{11} * 2^3 * v * 2^{11} * w * 2^5 * 2^5 * 2^{11} * 2^3 * 2^3 * u$$
$$* 2^{11} * v * 2^5 * 2^{11} * 2^3 * u * 2^{11} * w * 2^5 * 2^5 * 2^5)$$

(c) $\mathrm{Ax}_3(x)$: x is the Gödel number of an instance of axiom schema (A3):

$$(\exists u)_{u<x}(\exists v)_{v<x}(\mathrm{Fml}(u) \wedge \mathrm{Fml}(v)$$
$$\wedge x = 2^3 * 2^3 * 2^3 * 2^9 * v * 2^5 * 2^{11} * 2^3 * 2^9 * u * 2^5 * 2^5 * 2^{11}$$
$$* 2^3 * 2^3 * 2^3 * 2^9 * v * 2^5 * 2^{11} * u * 2^5 * 2^{11} * v * 2^5 * 2^5)$$

(d) $\mathrm{Ax}_4(x)$: x is the Gödel number of an instance of axiom schema (A4):

$$(\exists u)_{u<x}(\exists v)_{v<x}(\exists y)_{y<x}(\mathrm{Fml}(y) \wedge \mathrm{Trm}(u) \wedge \mathrm{EVbl}(2^v) \wedge \mathrm{Ff}(u, v, y)$$
$$\wedge x = 2^3 * 2^3 * 2^3 * 2^{13} * 2^v * 2^5 * y * 2^{11} * \mathrm{Sub}(y, u, v) * 2^5)$$

(e) $\mathrm{Ax}_5(x)$: x is the Gödel number of an instance of axiom schema (A5):

$$(\exists u)_{u<x}(\exists v)_{v<x}(\exists w)_{w<x}(\mathrm{Fml}(u) \wedge \mathrm{Fml}(w) \wedge \mathrm{EVbl}(2^v) \wedge \neg\mathrm{Fr}(u, v)$$
$$\wedge x = 2^3 * 2^3 * 2^3 * 2^{13} * 2^v * 2^5 * 2^3 * u * 2^{11} * w * 2^5 * 2^5$$
$$* 2^{11} * 2^3 * u * 2^{11} * 2^3 * 2^3 * 2^{13} * 2^v * 2^5 * w * 2^5 * 2^5 * 2^5)$$

(f) $\mathrm{LAX}(y)$: y is the Gödel number of a logical axiom of K

$$\mathrm{Ax}_1(y) \vee \mathrm{Ax}_2(y) \vee \mathrm{Ax}_3(y) \vee \mathrm{Ax}_4(y) \vee \mathrm{Ax}_5(y)$$

(14) The following *negation* function is primitive recursive. $\mathrm{Neg}(x)$: the Gödel number of $(\neg\mathscr{B})$ if x is the Gödel number of $\mathscr{B}$:

$$\mathrm{Neg}(x) = 2^3 * 2^9 * x * 2^5$$

(15) The following *conditional* function is primitive recursive. $\mathrm{Cond}(x, y)$: the Gödel number of $(\mathscr{B} \Rightarrow \mathscr{C})$ if x is the Gödel number of $\mathscr{B}$ and y is the Gödel number of $\mathscr{C}$:

$$\mathrm{Cond}(x, y) = 2^3 * x * 2^{11} * y * 2^5$$

(16) $\mathrm{Clos}(u)$: the Gödel number of the closure of $\mathscr{B}$ if u is the Gödel number of a wf $\mathscr{B}$. First, let $V(u) = \mu v_{v \leqslant u}(\mathrm{EVbl}(2^v) \wedge \mathrm{Fr}(u, v))$. V is primitive recursive (or recursive). $V(u)$ is the least Gödel number of a free variable of u (if there are any). Let $\mathrm{Sent}(u)$ be $\mathrm{Fml}(u) \wedge \neg(\exists v)_{v \leqslant u}\mathrm{Fr}(u, v)$. Sent is primitive recursive (or recursive). $\mathrm{Sent}(u)$ holds when and only when u is the Gödel number of a sentence (i.e., a closed wf). Now let

$$\mathrm{G}(u) = \begin{cases} 2^3 * 2^3 * 2^{13} * 2^{V(u)} * 2^5 * u * 2^5 & \text{if } \mathrm{Fml}(u) \wedge \neg\mathrm{Sent}(u) \\ u & \text{otherwise} \end{cases}$$

G is primitive recursive (or recursive). If u is the Gödel number of a wf $\mathscr{B}$ that is not a closed wf, then $G(u)$ is the Gödel number of $(\forall x)\mathscr{B}$, where x is the free variable of $\mathscr{B}$ that has the least Gödel number. Otherwise, $G(u) = u$. Now, let

$$H(u, 0) = G(u)$$
$$H(u, y+1) = G(H(u, y))$$

H is primitive recursive (or recursive). Finally,

$$\text{Clos}(u) = H(u, \mu y_{y \leqslant u}(H(u, y) = H(u, y+1)))$$

Thus, Clos is primitive recursive (or recursive).

PROPOSITION 3.27

Let K be a theory having a primitive recursive (or recursive) vocabulary and whose language contains the individual constant 0 and the function letter f_1^1 of $\mathscr{L}_A$. (Thus, all the numerals are terms of K. In particular, K can be S itself.) Then the following functions and relation are primitive recursive (or recursive).

(17) Num(y): the Gödel number of the expression $\bar{y}$

$$\text{Num}(0) = 2^{15}$$
$$\text{Num}(y+1) = 2^{49} * 2^3 * \text{Num}(y) * 2^5$$

Num is primitive recursive by virtue of the recursion rule (V).

(18) Nu(x): x is the Gödel number of a numeral

$$(\exists y)_{y<x}(x = \text{Num}(y))$$

Nu is primitive recursive by Proposition 3.18.

(19) D(u): the Gödel number of $\mathscr{B}(\bar{u})$, if u is the Gödel number of a wf $\mathscr{B}(x_1)$:

$$D(u) = \text{Sub}(u, \text{Num}(u), 21)$$

Thus, D is primitive recursive (or recursive). D is called the *diagonal function*.

DEFINITION

A theory K will be said to have a *primitive recursive* (or *recursive*) *axiom set* if the following property PrAx is primitive recursive (or recursive):

$$\text{PrAx}(y) : \ y \text{ is the Gödel number of a proper axiom of K}$$

Notice that S has a primitive recursive axiom set. Let $a_1, a_2, \ldots, a_8$ be the Gödel numbers of axioms (S1)- (S8). It is easy to see that a number y is the Gödel number of an instance of axiom schema (A9) if and only if

$$(\exists v)_{v<y}(\exists w)_{w<y}(\mathrm{EVbl}(2^v) \wedge \mathrm{Fml}(w)$$
$$\wedge y = 2^3 * \mathrm{Sub}(w, 2^{15}, v) * 2^{11} * 2^3 * 2^3 * 2^3 * 2^{13} * 2^v * 2^5$$
$$* 2^3 * w * 2^{11} * \mathrm{Sub}(w, 2^{49} * 2^3 * 2^v * 2^5, v) * 2^5 * 2^5 * 2^{11}$$
$$* 2^3 * 2^3 * 2^{13} * 2^v * 2^5 * w * 2^5 * 2^5)$$

Denote the displayed formula by $\mathrm{A}_9(y)$. Then y is the Gödel number of a proper axiom of S if and only if

$$y = a_1 \vee y = a_2 \vee \ldots \vee y = a_8 \vee \mathrm{A}_9(y)$$

Thus, $\mathrm{PrAx}(y)$ is primitive recursive for S.

PROPOSITION 3.28

Let K be a theory having a primitive recursive (or recursive) vocabulary and a primitive recursive (or recursive) axiom set. Then the following three relations are primitive recursive (or recursive).

(20) $\mathrm{Ax}(y)$: y is the Gödel number of an axiom of K:

$$\mathrm{LAX}(y) \vee \mathrm{PrAx}(y)$$

(21) $\mathrm{Prf}(y)$: y is the Gödel number of a proof in K:

$$(\exists u)_{u<y}(\exists v)_{v<y}(\exists z)_{z<y}(\exists w)_{w<y}([y = 2^w \wedge \mathrm{Ax}(w)] \vee$$
$$[\mathrm{Prf}(u) \wedge \mathrm{Fml}((u)_w) \wedge y = u * 2^v \wedge \mathrm{Gen}((u)_w, v)] \vee$$
$$[\mathrm{Prf}(u) \wedge \mathrm{Fml}((u)_z) \wedge \mathrm{Fml}((u)_w) \wedge y = u * 2^v \wedge \mathrm{MP}((u)_z, (u)_w, v)]$$
$$\vee [\mathrm{Prf}(u) \wedge y = u * 2^v \wedge \mathrm{Ax}(v)]$$

Apply Corollary 3.21.

(22) $\mathrm{Pf}(y, x)$: y is the Gödel number of a proof in K of the wf with Gödel number x

$$\mathrm{Prf}(y) \wedge x = (y)_{\ell h(y) \dot{-} 1}$$

The relations and functions of Propositions 3.26–3.28 should have the subscript 'K' attached to the corresponding signs to indicate the dependence on K. If we considered a different theory, then we would obtain different relations and functions.

Exercise

3.35

(a) If K is a theory for which the property $\mathrm{Fml}(y)$ is primitive recursive (or recursive), prove that K has a primitive recursive (or recursive) vocabulary.

(b) Let K be a theory for which the property $\mathrm{Ax}(y)$ is primitive recursive (or recursive).
(i) Show that K has a primitive recursive (or recursive) vocabulary.
(ii) Assuming also that no proper axiom of K is a logical axiom, prove that K has a primitive recursive (or recursive) axiom set.

PROPOSITION 3.29

Let K be a theory with equality whose language contains the individual constant 0 and the function letter f_1^1 and such that K has a primitive recursive (or recursive) vocabulary and axiom set. Also assume:

(*) For any natural numbers r and s, if $\vdash_{\mathrm{K}} \bar{r} = \bar{s}$, then $r = s$.

Then any function $f(x_1, \ldots, x_n)$ that is representable in K is recursive.

Proof

Let $\mathscr{B}(x_1, \ldots, x_n, x_{n+1})$ be a wf of K that represents f. Let $P_{\mathscr{B}}(u_1, \ldots, u_n, u_{n+1}, y)$ mean that y is the Gödel number of a proof in K of the wf $\mathscr{B}(\bar{u}_1, \ldots, \bar{u}_n, \bar{u}_{n+1})$. Note that, if $P_{\mathscr{B}}(u_1, \ldots, u_n, u_{n+1}, y)$, then $f(u_1, \ldots, u_n) = u_{n+1}$. (In fact, let $f(u_1, \ldots, u_n) = r$. Since $\mathscr{B}$ represents f in K, $\vdash_{\mathrm{K}} \mathscr{B}(\bar{u}_1, \ldots, u_n, \bar{r})$ and $\vdash_{\mathrm{K}} (\exists_1 y)\mathscr{B}(\bar{u}_1, \ldots, \bar{u}_n, y)$. By hypothesis, $P_{\mathscr{B}}(u_1, \ldots, u_n, u_{n+1}, y)$. Hence, $\vdash_{\mathrm{K}} \mathscr{B}(\bar{u}_1, \ldots, \bar{u}_n, \bar{u}_{n+1})$. Since K is a theory with equality, it follows that $\vdash_{\mathrm{K}} \bar{r} = \bar{u}_{n+1}$. By (*), $r = u_{n+1}$.) Now let m be the Gödel number of $\mathscr{B}(x_1, \ldots, x_n, x_{n+1})$. Then $P_{\mathscr{B}}(u_1, \ldots, u_n, u_{n+1}, y)$ is equivalent to:

$$\mathrm{Pf}(y, \mathrm{Sub}(\ldots \mathrm{Sub}(\mathrm{Sub}(\mathrm{m}, \mathrm{Num}(u_1), 21), \mathrm{Num}(u_2), 29) \ldots \mathrm{Num}(u_{n+1}), 21 + 8n))$$

So, by Propositions 3.26–3.28, $P_{\mathscr{B}}(u_1, \ldots, u_n, u_{n+1}, y)$ is primitive recursive (or recursive). Now consider any natural numbers $k_1, \ldots, k_n$. Let $f(k_1, \ldots, k_n) = r$. Then $\vdash_{\mathrm{K}} \mathscr{B}(\bar{k}_1, \ldots, \bar{k}_n, \bar{r})$. Let j be the Gödel number of a proof in K of $\mathscr{B}(\bar{k}_1, \ldots, \bar{k}_n, \bar{r})$. Then $P_{\mathscr{B}}(k_1, \ldots, k_n, r, j)$. Thus, for any $x_1, \ldots, x_n$, there is some y such that $P_{\mathscr{B}}(x_1, \ldots, x_n, (y)_0, (y)_1)$. Then, by Exercise 3.16(c), $\mu y(P_{\mathscr{B}}(x_1, \ldots, x_n, (y)_0, (y)_1))$ is recursive. But, $f(x_1, \ldots, x_n) = (\mu y(P_{\mathscr{B}}(x_1, \ldots, x_n, (y)_0, (y)_1)))_0$ and, therefore, f is recursive.

Exercise

3.36 Let K be a theory whose language contains the predicate letter $=$, the individual constant 0, and the function letter f_1^1.
(a) If K satisfies hypothesis (*) of Proposition 3.29, prove that K must be consistent.

(b) If K is inconsistent, prove that every number-theoretic function is representable in K.
(c) If K is consistent and the identity relation $x = y$ is expressible in K, show that K satisfies hypothesis (*) of Proposition 3.29.

COROLLARY 3.30

Assume S consistent. Then the class of recursive functions is identical with the class of functions representable in S.

Proof

We have observed that S has a primitive recursive vocabulary and axiom set. By Exercise 3.36(c) and the already noted fact that the identity relation is expressible in S, we see that Proposition 3.29 entails that every function representable in S is recursive. On the other hand, Proposition 3.24 tells us that every recursive function is representable in S.

In Chapter 5, it will be made plausible that the notion of recursive function is a precise mathematical equivalent of the intuitive idea of *effectively computable function.*

COROLLARY 3.31

A number-theoretic relation $R(x_1, \ldots, x_n)$ is recursive if and only if it is expressible in S.

Proof

By definition, R is recursive if and only if C_R is recursive. By Corollary 3.30, C_R is recursive if and only if C_R is representable in S. But, by Proposition 3.13, C_R is representable in S if and only if R is expressible in S.

It will be helpful later to find weaker theories than S for which the representable functions are identical with the recursive functions. Analysis of the proof of Proposition 3.24 leads us to the following theory.

Robinson's System

Consider the theory in the language $\mathscr{L}_A$ with the following finite list of proper axioms.

(1) $x_1 = x_1$
(2) $x_1 = x_2 \Rightarrow x_2 = x_1$

(3) $x_1 = x_2 \Rightarrow (x_2 = x_3 \Rightarrow x_1 = x_3)$
(4) $x_1 = x_2 \Rightarrow x_1' = x_2'$
(5) $x_1 = x_2 \Rightarrow (x_1 + x_3 = x_2 + x_3 \wedge x_3 + x_1 = x_3 + x_2)$
(6) $x_1 = x_2 \Rightarrow (x_1 \cdot x_3 = x_2 \cdot x_3 \wedge x_3 \cdot x_1 = x_3 \cdot x_2)$
(7) $x_1' = x_2' \Rightarrow x_1 = x_2$
(8) $0 \neq x_1'$
(9) $x_1 \neq 0 \Rightarrow (\exists x_2)(x_1 = x_2')$
(10) $x_1 + 0 = x_1$
(11) $x_1 + x_2' = (x_1 + x_2)'$
(12) $x_1 \cdot 0 = 0$
(13) $x_1 \cdot x_2' = (x_1 \cdot x_2) + x_1$
(14) $(x_2 = x_1 \cdot x_3 + x_4 \wedge x_4 < x_1 \wedge x_2 = x_1 \cdot x_5 + x_6 \wedge x_6 < x_1) \Rightarrow$
$x_4 = x_6$ (uniqueness of remainder)

We shall call this theory RR. Clearly, RR is a subtheory of S, since all the axioms of RR are theorems of S. In addition, it follows from Proposition 2.25 and axioms (1)–(6) that RR is a theory with equality. (The system Q of axioms (1)–(13) is due to R.M. Robinson (1950). Axiom (14) has been added to make one of the proofs below easier.) Notice that RR has only a finite number of proper axioms.

LEMMA 3.32

In RR, the following are theorems.

(a) $\bar{n} + \bar{m} = \overline{n+m}$ for any natural numbers n and m
(b) $\bar{n} \cdot \bar{m} = \overline{n \cdot m}$ for any natural numbers n and m
(c) $\bar{n} \neq \bar{m}$ for any natural numbers such that $n \neq m$
(d) $\bar{n} < \bar{m}$ for any natural numbers n and m such that $n < m$
(e) $x \not< 0$
(f) $x \leqslant \bar{n} \Rightarrow x = 0 \vee x = \bar{1} \vee \ldots \vee x = \bar{n}$ for any natural number n
(g) $x \leqslant \bar{n} \vee \bar{n} \leqslant x$ for any natural number n

Proof

Parts (a)–(c) are proved the same way as Proposition 3.6(a). Parts (d)–(g) are left as exercises.

PROPOSITION 3.33

All recursive functions are representable in RR.

Proof

The initial functions Z, N, and U_i^n are representable in RR by the same wfs as in Examples 1–3, page172. That the substitution rule does not lead out of the class of functions representable in RR is proved in the same way as in Example 4 on page 172. For the recursion rule, first notice that the proof of Lemma 3.22 is a demonstration that Gödel's beta function $\beta(x_1, x_2, x_3)$ is strongly representable in RR. (Axiom (14) is used for the uniqueness part.) Now, a careful examination of the treatment of the recursion rule in the proof of Proposition 3.24 reveals that all the required theorems are theorems of RR. The argument given for the restricted μ-operator rule also remains valid for RR.

By Proposition 3.33, all recursive functions are representable in any extension of RR. Hence, by Proposition 3.29 and Exercise 3.36(c), in any consistent extension of RR in the language $\mathscr{L}_A$ that has a recursive axiom set, the class of representable functions is the same as the class of recursive functions. Moreover, by Proposition 3.13, the relations expressible in such a theory are the recursive relations.

Exercises

3.37^{D} Show that RR is a proper subtheory of S. [*Hint*: Find a model for RR that is not a model for S.] (*Remark*: Not only is S different from RR, but it is not finitely axiomatizable at all, that is, there is no theory K having only a finite number of proper axioms, whose theorems are the same as those of S. This was proved by Ryll-Nardzewski (1953).)

3.38 Show that axiom (14) of RR is not provable from axioms (1)–(13) and, therefore, that Q is a proper subtheory of RR. [*Hint*: Find a model of (1)–(13) for which (14) is not true.]

3.39 Let K be a theory in the language $\mathscr{L}_A$ with just one proper axiom: $(\forall x_1)(\forall x_2)x_1 = x_2$.

(a) Show that K is a consistent theory with equality.
(b) Prove that all number-theoretic functions are representable in K.
(c) Which number-theoretic relations are expressible in K? [*Hint*: Use elimination of quantifiers.]
(d) Show that the hypothesis $\vdash_K 0 \neq 1$ cannot be eliminated from Proposition 3.13.
(e) Show that, in Proposition 3.29, the hypothesis $(*)$ cannot be replaced by the assumption that K is consistent.

3.40 Let R be the theory in the language $\mathscr{L}_A$ having as proper axioms the equality axioms (1)–(6) of RR as well as the following five axiom schemas, in which n and m are arbitrary natural numbers:

(R1) $\bar{n} + \bar{m} = \overline{n+m}$
(R2) $\bar{n} \cdot \bar{m} = \overline{n \cdot m}$

(R3) $\bar{n} \neq \bar{m}$ if $n \neq m$
(R4) $x \leqslant \bar{n} \Rightarrow x = 0 \vee \ldots \vee x = \bar{n}$
(R5) $x \leqslant \bar{n} \vee \bar{n} \leqslant x$

Prove the following.

(a) R is not finitely axiomatizable. [*Hint*: Show that every finite subset of the axioms of R has a model that is not a model of R.]
(b) R is a proper subtheory of Q.
(c)[D] Every recursive function is representable in R. (See Monk, 1976, p. 248.)
(d) The functions representable in R are the recursive functions.
(e) The relations expressible in R are the recursive relations.

3.5 THE FIXED-POINT THEOREM. GÖDEL'S INCOMPLETENESS THEOREM

If K is a theory in the language $\mathscr{L}_A$, recall that the diagonal function D has the property that, if u is the Gödel number of a wf $\mathscr{B}(x_1)$, then $D(u)$ is the Gödel number of the wf $\mathscr{B}(\bar{u})$.

NOTATION

When $\mathscr{C}$ is an expression of a theory and the Gödel number of $\mathscr{C}$ is q, then we shall denote the numeral $\bar{q}$ by $\ulcorner \mathscr{C} \urcorner$. We can think of $\ulcorner \mathscr{C} \urcorner$ as being a 'name' for $\mathscr{C}$ within the language $\mathscr{L}_A$.

PROPOSITION 3.34 (DIAGONALIZATION LEMMA)

Assume that the diagonal function D is representable in a theory with equality K in the language $\mathscr{L}_A$. Then, for any wf $\mathscr{E}(x_1)$ in which x_1 is the only free variable, there exists a closed wf $\mathscr{C}$ such that

$$\vdash_K \mathscr{C} \Leftrightarrow \mathscr{E}(\ulcorner \mathscr{C} \urcorner)$$

Proof

Let $\mathscr{D}(x_1, x_2)$ be a wf representing D in K. Construct the wf

$$(\nabla) \qquad (\forall x_2)(\mathscr{D}(x_1, x_2) \Rightarrow \mathscr{E}(x_2))$$

Let m be the Gödel number of (∇). Now substitute $\bar{m}$ for x_1 in (∇):

$$(\mathscr{C}) \qquad (\forall x_2)(\mathscr{D}(\bar{m}, x_2) \Rightarrow \mathscr{E}(x_2))$$

Let q be the Gödel number of this wf $\mathscr{C}$. So, $\bar{q}$ is $\ulcorner \mathscr{C} \urcorner$. Clearly, $D(m) = q$. (In fact, m is the Gödel number of a wf $\mathscr{B}(x_1)$, namely, (∇), and q is the Gödel number of $\mathscr{B}(\bar{m})$.) Since $\mathscr{D}$ represents D in K,

$$(\partial) \qquad \vdash_{\mathrm{K}} \mathscr{D}(\bar{m}, \bar{q})$$

(a) Let us show $\vdash_{\mathrm{K}} \mathscr{C} \Rightarrow \mathscr{E}(\bar{q})$.

1. $\mathscr{C}$	Hyp
2. $(\forall x_2)(\mathscr{D}(\bar{m}, x_2) \Rightarrow \mathscr{E}(x_2))$	Same as 1
3. $\mathscr{D}(\bar{m}, \bar{q}) \Rightarrow \mathscr{E}(\bar{q}))$	2, rule A4
4. $\mathscr{D}(\bar{m}, \bar{q})$	(∂)
5. $\mathscr{E}(\bar{q})$	3, 4, MP
6. $\mathscr{C} \vdash_{\mathrm{K}} \mathscr{E}(\bar{q})$	1–5
7. $\vdash_{\mathrm{K}} \mathscr{C} \Rightarrow \mathscr{E}(\bar{q})$	1–6, Corollary 2.6

(b) Let us prove $\vdash_{\mathrm{K}} \mathscr{E}(\bar{q}) \Rightarrow \mathscr{C}$.

1. $\mathscr{E}(\bar{q})$	Hyp
2. $\mathscr{D}(\bar{m}, x_2)$	Hyp
3. $(\exists_1 x_2)\mathscr{D}(\bar{m}, x_2)$	$\mathscr{D}$ represents D
4. $\mathscr{D}(\bar{m}, \bar{q})$	(∂)
5. $x_2 = \bar{q}$	2–4, properties of $=$
6. $\mathscr{E}(x_2)$	1, 5, substitutivity of $=$
7. $\mathscr{E}(\bar{q}), \mathscr{D}(\bar{m}, x_2) \vdash_{\mathrm{K}} \mathscr{E}(x_2)$	1–6
8. $\mathscr{E}(\bar{q}) \vdash_{\mathrm{K}} \mathscr{D}(\bar{m}, x_2) \Rightarrow \mathscr{E}(x_2)$	1–7, Corollary 2.6
9. $\mathscr{E}(\bar{q}) \vdash_{\mathrm{K}} (\forall x_2)(\mathscr{D}(\bar{m}, x_2) \Rightarrow \mathscr{E}(x_2))$	8, Gen
10. $\vdash_{\mathrm{K}} \mathscr{E}(\bar{q}) \Rightarrow (\forall x_2)(\mathscr{D}(\bar{m}, x_2) \Rightarrow \mathscr{E}(x_2))$	1–9, Corollary 2.6
11. $\vdash_{\mathrm{K}} \mathscr{E}(\bar{q}) \Rightarrow \mathscr{C}$	Same as 10

From parts (a) and (b), by biconditional introduction, $\vdash_{\mathrm{K}} \mathscr{C} \Leftrightarrow \mathscr{E}(\bar{q})$.

PROPOSITION 3.35 (FIXED-POINT THEOREM)†

Assume that all recursive functions are representable in a theory with equality K in the language $\mathscr{L}_A$. Then, for any wf $\mathscr{E}(x_1)$ in which x_1 is the only free variable, there is a closed wf $\mathscr{C}$ such that

$$\vdash_{\mathrm{K}} \mathscr{C} \Leftrightarrow \mathscr{E}(\ulcorner \mathscr{C} \urcorner)$$

†The terms 'fixed-point theorem' and 'diagonalization lemma' are often used interchangeably, but I have adopted the present terminology for convenience of reference. The central idea seems to have first received explicit mention by Carnap (1934), who pointed out that the result was implicit in the work of Gödel (1931). The use of indirect self-reference was the key idea in the explosion of progress in mathematical logic that began in the 1930s

Proof

By Proposition 3.27, D is recursive.† Hence, D is representable in K and Proposition 3.34 is applicable.

By Proposition 3.33, the fixed-point theorem holds when K is RR or any extension of RR. In particular, it holds for S.

DEFINITIONS

Let K be any theory whose language contains the individual constant 0 and the function letter f_1^1. Then K is said to be *ω-consistent* if, for every wf $\mathscr{B}(x)$ of K containing x as its only free variable, if $\vdash_K \neg\mathscr{B}(\bar{n})$ for every natural number n, then it is not the case that $\vdash_K (\exists x)\mathscr{B}(x)$.

Let K be any theory in the language $\mathscr{L}_A$. K is said to be a *true theory* if all proper axioms of K are true in the standard model. (Since all logical axioms are true in all models and MP and Gen lead from wfs true in a model to wfs true in that model, all theorems of a true theory will be true in the standard model.)

Any true theory K must be ω-consistent. (In fact, if $\vdash_K \neg\mathscr{B}(\bar{n})$ for all natural numbers n, then $\mathscr{B}(x)$ is false for every natural number and, therefore, $(\exists x)\mathscr{B}(x)$ cannot be true for the standard model. Hence, $(\exists x)\mathscr{B}(x)$ cannot be a theorem of K.) In particular, RR and S are ω-consistent.

PROPOSITION 3.36

If K is ω-consistent, then K is consistent.

Proof

Let $\mathscr{E}(x)$ be any wf containing x as its only free variable. Let $\mathscr{B}(x)$ be $\mathscr{E}(x) \wedge \neg\mathscr{E}(x)$. Then $\neg\mathscr{B}(\bar{n})$ is an instance of a tautology. Hence, $\vdash_K \neg\mathscr{B}(\bar{n})$ for every natural number n. By ω-consistency, not-$\vdash_K (\exists x)\mathscr{B}(x)$. Therefore, K is consistent. (Remember that *every* wf is provable in an inconsistent

†In fact, D is primitive recursive, since K, being a theory in $\mathscr{L}_A$, has a primitive recursive vocabulary.

theory, by virtue of the tautology $\neg A \Rightarrow (A \Rightarrow B)$. Hence, if at least one wf is not provable, the theory must be consistent.)

It will turn out later that the converse of Proposition 3.36 does not hold.

DEFINITION

An *undecidable sentence* of a theory K is a closed wf $\mathscr{B}$ of K such that neither $\mathscr{B}$ nor $\neg\mathscr{B}$ is a theorem of K, that is, such that not-$\vdash_K \mathscr{B}$ and not-$\vdash_K \neg\mathscr{B}$.

Gödel's incompleteness theorem

Let K be a theory with equality in the language $\mathscr{L}_A$ satisfying the following three conditions:

1. K has a recursive axiom set (that is, PrAx(y) is recursive).
2. $\vdash_K 0 \neq \bar{1}$.
3. Every recursive function is representable in K.

By assumption 1, Propositions 3.26–3.28 are applicable. By assumptions 2 and 3 and Proposition 3.13, every recursive relation is expressible in K. By assumption 3, the fixed-point theorem is applicable. Note that K can be taken to be RR, S, or, more generally, any extension of RR having a recursive axiom set. Recall that Pf(y,x) means that y is the Gödel number of a proof in K of a wf with Gödel number x. By Proposition 3.28, Pf is recursive. Hence, Pf is expressible in K by a wf $\mathscr{P}f(x_2, x_1)$. Let $\mathscr{E}(x_1)$ be the wf $(\forall x_2)\neg\mathscr{P}f(x_2, x_1)$. By the fixed-point theorem, there must be a closed wf $\mathscr{G}$ such that

$$(\$) \qquad \vdash_K \mathscr{G} \Leftrightarrow (\forall x_2)\neg\mathscr{P}f(x_2, \ulcorner\mathscr{G}\urcorner)$$

Observe that, in terms of the standard interpretation, $(\forall x_2)\neg\mathscr{P}f(x_2, \ulcorner\mathscr{G}\urcorner)$ says that there is no natural number that is the Gödel number of a proof in K of the wf $\mathscr{G}$, which is equivalent to asserting that there is no proof in K of $\mathscr{G}$. Hence, $\mathscr{G}$ is equivalent in K to an assertion that $\mathscr{G}$ is unprovable in K. In other words, $\mathscr{G}$ says 'I am not provable in K'. This is an analogue of the liar paradox: 'I am lying' (that is, 'I am not true'). However, although the liar paradox leads to a contradiction, Gödel (1931) showed that $\mathscr{G}$ is an undecidable sentence of K. We shall refer to $\mathscr{G}$ as a *Gödel sentence* for K.

PROPOSITION 3.37 (GÖDEL'S INCOMPLETENESS THEOREM)

Let K satisfy conditions 1–3. Then:
(a) If K is consistent, not-$\vdash_K \mathscr{G}$.

(b) If K is ω-consistent, not-$\vdash_K \neg\mathscr{G}$.
Hence, if K is ω-consistent, $\mathscr{G}$ is an undecidable sentence of K.

Proof

Let q be the Gödel number of $\mathscr{G}$.

(a) Assume $\vdash_K \mathscr{G}$. Let r be the Gödel number of a proof in K of $\mathscr{G}$. Then Pf(r, q). Hence, $\vdash_K \mathscr{Pf}(\bar{r}, \bar{q})$, that is $\vdash_K \mathscr{Pf}(\bar{r}, \ulcorner\mathscr{G}\urcorner)$. But, from (\$) above by biconditional elimination, $\vdash_K (\forall x_2)\neg\mathscr{Pf}(x_2, \ulcorner\mathscr{G}\urcorner)$. By rule A4, $\vdash_K \neg\mathscr{Pf}(\bar{r}, \ulcorner\mathscr{G}\urcorner)$. Therefore, K is inconsistent.

(b) Assume K is ω-consistent and $\vdash_K \neg\mathscr{G}$. From (\$) by biconditional elimination, $\vdash_K \neg(\forall x_2)\neg\mathscr{Pf}(x_2, \ulcorner\mathscr{G}\urcorner)$, which abbreviates to

$$(*) \qquad \vdash_K (\exists x_2)\mathscr{Pf}(x_2, \ulcorner\mathscr{G}\urcorner)$$

On the other hand, since K is ω-consistent, Proposition 3.36 implies that K is consistent. But, $\vdash_K \neg\mathscr{G}$. Hence, not-$\vdash_K \mathscr{G}$, that is, there is no proof in K of $\mathscr{G}$. So, Pf(n, q) is false for every natural number n and, therefore, $\vdash_K \neg\mathscr{Pf}(\bar{n}, \ulcorner\mathscr{G}\urcorner)$ for every natural number n. (Remember that $\ulcorner\mathscr{G}\urcorner$ is $\bar{q}$.) By ω-consistency, not-$\vdash_K (\exists x_2)\mathscr{Pf}(x_2, \ulcorner\mathscr{G}\urcorner)$, contradicting $(*)$.

REMARKS

Gödel's incompleteness theorem has been established for any theory with equality K in the language $\mathscr{L}_A$ that satisfies conditions 1–3 above. Assume that K also satisfies the following condition:

$$(+) \qquad \text{K is a true theory.}$$

(In particular, K can be S or any subtheory of S.) Proposition 3.37(a) shows that, if K is consistent, $\mathscr{G}$ is not provable in K. But, under the standard interpretation, $\mathscr{G}$ asserts its own unprovability in K. Therefore, $\mathscr{G}$ *is true for the standard interpretation.*

Moreover, when K is a true theory, the following simple intuitive argument can be given for the undecidability of $\mathscr{G}$ in K.

(i) Assume $\vdash_K \mathscr{G}$. Since $\vdash_K \mathscr{G} \Leftrightarrow (\forall x_2)\neg\mathscr{Pf}(x_2, \ulcorner\mathscr{G}\urcorner)$, it follows that $\vdash_K (\forall x_2)\neg\mathscr{Pf}(x_2, \ulcorner\mathscr{G}\urcorner)$. Since K is a true theory, $(\forall x_2)\neg\mathscr{Pf}(x_2, \ulcorner\mathscr{G}\urcorner)$ is true for the standard interpretation. But this wf says that $\mathscr{G}$ is not provable in K, contradicting our original assumption. Hence, not-$\vdash_K \mathscr{G}$.

(ii) Assume $\vdash_K \neg\mathscr{G}$. Since $\vdash_K \mathscr{G} \Leftrightarrow (\forall x_2)\neg\mathscr{Pf}(x_2, \ulcorner\mathscr{G}\urcorner)$, $\vdash_K \neg(\forall x_2)\neg\mathscr{Pf}(x_2, \ulcorner\mathscr{G}\urcorner)$. So, $\vdash_K (\exists x_2)\mathscr{Pf}(x_2, \ulcorner\mathscr{G}\urcorner)$. Since K is a true theory, this wf is true for the standard interpretation, that is, $\mathscr{G}$ is provable in K. This contradicts the result of (i). Hence, not-$\vdash_K \neg\mathscr{G}$.

Exercises

3.41 Let $\mathscr{G}$ be a Gödel sentence for S. Let S_g be the extension of S obtained by adding $\neg\mathscr{G}$ as a new axiom. Prove that, if S is consistent, then S_g is consistent, but not ω-consistent.

3.42 A theory K whose language has the individual constant 0 and function letter f_1^1 is said to be *ω-incomplete* if there is a wf $\mathscr{E}(x)$ with one free variable x such that $\vdash_K \mathscr{E}(\bar{n})$ for every natural number n, but it is not the case that $\vdash_K (\forall x)\mathscr{E}(x)$. If K is a consistent theory with equality in the language $\mathscr{L}_A$ and satisfies conditions 1–3 on page 206, show that K is ω-incomplete. (In particular, RR and S are ω-incomplete.)

3.43 Let K be a theory whose language contains the individual constant 0 and function letter f_1^1. Show that, if K is a consistent and ω-inconsistent, then K is ω-incomplete.

3.44 Prove that S, as well as any consistent extension of S having a recursive axiom set, is not a scapegoat theory. (See page 87.)

3.45 Show that there is an ω-consistent extension K of S such that K is not a true theory. [*Hint*: Use the fixed point theorem.]

The Gödel–Rosser incompleteness theorem

The proof of undecidability of a Gödel sentence $\mathscr{G}$ required the assumption of ω-consistency. We will now prove a result of Rosser (1936) showing that, at the cost of a slight increase in the complexity of the undecidable sentence, the assumption of ω-consistency can be replaced by consistency.

As before, let K be a theory with equality in the language $\mathscr{L}_A$ satisfying conditions 1–3 on page 206. In addition, assume:

4. $\vdash_K x \leqslant \bar{n} \Rightarrow x = 0 \vee x = \bar{1} \vee \ldots \vee x = \bar{n}$ for every natural number n.
5. $\vdash_K x \leqslant \bar{n} \vee \bar{n} \leqslant x$ for every natural number n.

Thus, K can be any extension of RR with a recursive axiom set. In particular, K can be RR or S.

Recall that, by Proposition 3.26 (14), Neg is a primitive recursive function such that, if x is the Gödel number of a wf $\mathscr{B}$, then $\mathrm{Neg}(x)$ is the Gödel number of $(\neg\mathscr{B})$. Since all recursive functions are representable in K, let $\mathscr{N}eg(x_1, x_2)$ be a wf that represents Neg in K. Now construct the following wf $\mathscr{E}(x_1)$:

$$(\forall x_2)(\mathscr{P}f(x_2, x_1) \Rightarrow (\forall x_3)(\mathscr{N}eg(x_1, x_3) \Rightarrow (\exists x_4)(x_4 \leqslant x_2 \wedge \mathscr{P}f(x_4, x_3))))$$

By the fixed-point theorem, there is a closed wf $\mathscr{R}$ such that

$$(*) \qquad \vdash_K \mathscr{R} \Leftrightarrow \mathscr{E}(\ulcorner \mathscr{R} \urcorner)$$

$\mathscr{R}$ is called a *Rosser sentence* for K. Notice what the intuitive meaning of $\mathscr{R}$ is under the standard interpretation. $\mathscr{R}$ asserts that, if $\mathscr{R}$ has a proof in K,

say with Gödel number x_2, then $\neg\mathscr{R}$ has a proof in K with Gödel number smaller than x_2. This is a roundabout way for $\mathscr{R}$ to claim its own unprovability under the assumption of the consistency of K.

PROPOSITION 3.38 (GÖDEL–ROSSER THEOREM)

Let K satisfy conditions 1–5. If K is consistent, then $\mathscr{R}$ is an undecidable sentence of K.

Proof

Let p be the Gödel number of $\mathscr{R}$. Thus, $\ulcorner\mathscr{R}\urcorner$ is $\bar{p}$. Let j be the Gödel number of $\neg\mathscr{R}$.

(a) Assume $\vdash_K \mathscr{R}$. Since $\vdash_K \mathscr{R} \Leftrightarrow \mathscr{E}(\ulcorner\mathscr{R}\urcorner)$, biconditional elimination yields $\vdash_K \mathscr{E}(\ulcorner\mathscr{R}\urcorner)$, that is:

$$\vdash_K (\forall x_2)(\mathscr{P}\!f(x_2,\bar{p}) \Rightarrow (\forall x_3)(\mathscr{N}eg(\bar{p},x_3) \Rightarrow (\exists x_4)(x_4 \leqslant x_2 \wedge \mathscr{P}\!f(x_4,x_3))))$$

Let k be the Gödel number of a proof in K of $\mathscr{R}$. Then $\mathrm{Pf}(k,p)$ and, therefore, $\vdash_K \mathscr{P}\!f(\bar{k},\bar{p})$. Applying rule A4 to $\mathscr{E}(\ulcorner\mathscr{R}\urcorner)$, we obtain

$$\vdash_K \mathscr{P}\!f(\bar{k},\bar{p}) \Rightarrow (\forall x_3)(\mathscr{N}eg(\bar{p},x_3) \Rightarrow (\exists x_4)(x_4 \leqslant \bar{k} \wedge \mathscr{P}\!f(x_4,x_3)))$$

So, by MP,

$$(\%) \qquad \vdash_K (\forall x_3)(\mathscr{N}eg(\bar{p},x_3) \Rightarrow (\exists x_4)(x_4 \leqslant \bar{k} \wedge \mathscr{P}\!f(x_4,x_3)))$$

Since j is the Gödel number of $\neg\mathscr{R}$, we have $\mathrm{Neg}(p,j)$, and, therefore, $\vdash_K \mathscr{N}eg(\bar{p},\bar{j})$. Applying rule A4 to (%), we obtain $\vdash_K \mathscr{N}eg(\bar{p},\bar{j}) \Rightarrow (\exists x_4)(x_4 \leqslant \bar{k} \wedge \mathscr{P}\!f(x_4,\bar{j}))$. Hence, by MP, $\vdash_K (\exists x_4)(x_4 \leqslant \bar{k} \wedge \mathscr{P}\!f(x_4,\bar{j}))$, which is an abbreviation for

$$(\#) \qquad \vdash_K \neg(\forall x_4)\neg(x_4 \leqslant \bar{k} \wedge \mathscr{P}\!f(x_4,\bar{j}))$$

Since $\vdash_K \mathscr{R}$, the consistency of K implies not-$\vdash_K \neg\mathscr{R}$. Hence, $\mathrm{Pf}(n,j)$ is false for all natural numbers n. Therefore, $\vdash_K \neg\mathscr{P}\!f(\bar{n},\bar{j})$ for all natural numbers n. Since K is a theory with equality, $\vdash_K x_4 = \bar{n} \Rightarrow \neg\mathscr{P}\!f(x_4,\bar{j})$ for all natural numbers n. By condition 4,

$$(\phi) \qquad \vdash_K x_4 \leqslant \bar{k} \Rightarrow x_4 = 0 \vee x_4 = \bar{1} \vee \ldots \vee x_4 = \bar{k}$$

But

$$(\phi\,\phi) \qquad \vdash_K x_4 = \bar{n} \Rightarrow \neg\mathscr{P}\!f(x_4,\bar{j}) \text{ for } n = 0, 1, \ldots, k$$

So, by a suitable tautology, (ϕ) and $(\phi\,\phi)$ yield $\vdash_K x_4 \leqslant \bar{k} \Rightarrow \neg\mathscr{P}\!f(x_4,\bar{j})$ and then, by another tautology, $\vdash_K \neg(x_4 \leqslant \bar{k} \wedge \mathscr{P}\!f(x_4,\bar{j}))$. By Gen,

$\vdash_K (\forall x_4)\neg(x_4 \leqslant \bar{k} \wedge \mathscr{P}\!f(x_4, \bar{j}))$. This, together with (#), contradicts the consistency of K.

(b) Assume $\vdash_K \neg\mathscr{R}$. Let m be the Gödel number of a proof of $\neg\mathscr{R}$ in K. So, Pf(m, j) is true and, therefore, $\vdash_K \mathscr{P}\!f(\bar{m}, \bar{j})$. Hence, by an application of rule E4 and the deduction theorem, $\vdash_K \bar{m} \leqslant x_2 \Rightarrow (\exists x_4)(x_4 \leqslant x_2 \wedge \mathscr{P}\!f(x_4, \bar{j}))$. By consistency of K, not-$\vdash_K \mathscr{R}$ and, therefore, Pf(n, p) is false for all natural numbers n. Hence, $\vdash_K \neg\mathscr{P}\!f(\bar{n}, \bar{p})$ for all natural numbers n. By condition 4, $\vdash_K x_2 \leqslant \bar{m} \Rightarrow x_2 = 0 \vee x_2 = \bar{1} \vee \ldots \vee x_2 = \bar{m}$. Hence, $\vdash_K x_2 \leqslant \bar{m} \Rightarrow \neg\mathscr{P}\!f\ (x_2, \bar{p})$. Consider the following derivation.

1. $\mathscr{P}\!f(x_2, \bar{p})$	Hyp
2. $\mathscr{N}eg(\bar{p}, x_3)$	Hyp
3. $x_2 \leqslant \bar{m} \vee \bar{m} \leqslant x_2$	Condition 5
4. $\bar{m} \leqslant x_2 \Rightarrow (\exists x_4)(x_4 \leqslant x_2 \wedge \mathscr{P}\!f(x_4, \bar{j}))$	Proved above
5. $x_2 \leqslant \bar{m} \Rightarrow \neg\mathscr{P}\!f(x_2, p)$	Proved above
6. $\neg\mathscr{P}\!f(x_2, \bar{p}) \vee (\exists x_4)(x_4 \leqslant x_2 \wedge \mathscr{P}\!f(x_4, \bar{j}))$	3–5, tautology
7. $(\exists x_4)(x_4 \leqslant x_2 \wedge \mathscr{P}\!f(x_4, \bar{j}))$	1, 6, disjunction rule
8. $\mathscr{N}eg(\bar{p}, \bar{j})$	Proved in part (a)
9. $(\exists_1 x_3)\mathscr{N}eg(\bar{p}, x_3)$	$\mathscr{N}eg$ represents Neg
10. $x_3 = \bar{j}$	2, 8, 9, properties of $=$
11. $(\exists x_4)(x_4 \leqslant x_2 \wedge \mathscr{P}\!f(x_4, x_3))$	7, 10, substitutivity of $=$
12. $\mathscr{P}\!f(x_2, \bar{p}), \mathscr{N}eg(\bar{p}, x_3) \vdash_K (\exists x_4)$ $(x_4 \leqslant x_2 \wedge \mathscr{P}\!f(x_4, x_3))$	1–11
13. $\mathscr{P}\!f(x_2, \bar{p}) \vdash_K \mathscr{N}eg(\bar{p}, x_3)$ $\Rightarrow (\exists x_4)(x_4 \leqslant x_2 \wedge \mathscr{P}\!f(x_4, x_3))$	1–12, Corollary 2.6
14. $\mathscr{P}\!f(x_2, \bar{p}) \vdash_K (\forall x_3)(\mathscr{N}eg(\bar{p}, x_3)$ $\Rightarrow (\exists x_4)(x_4 \leqslant x_2 \wedge \mathscr{P}\!f(x_4, x_3)))$	13, Gen
15. $\vdash_K \mathscr{P}\!f(x_2, \bar{p}) \Rightarrow (\forall x_3)(\mathscr{N}eg(\bar{p}, x_3)$ $\Rightarrow (\exists x_4)(x_4 \leqslant x_2 \wedge \mathscr{P}\!f(x_4, x_3)))$	1–14, Corollary 2.6
16. $\vdash_K (\forall x_2)(\mathscr{P}\!f(x_2, \bar{p}) \Rightarrow (\forall x_3)$ $(\mathscr{N}eg(\bar{p}, x_3) \Rightarrow (\exists x_4)(x_4 \leqslant x_2 \wedge$ $\mathscr{P}\!f(x_4, x_3))))$	15, Gen
17. $\vdash_K \mathscr{R}$	(*1, biconditional elimination)

Thus, $\vdash_K \mathscr{R}$ and $\vdash_K \neg\mathscr{R}$, contradicting the consistency of K.

The Gödel and Rosser sentences for the theory S are undecidable sentences of S. They have a certain intuitive metamathematical meaning; for example, a Gödel sentence $\mathscr{G}$ asserts that $\mathscr{G}$ is unprovable in S. Until recently, no undecidable sentences of S were known that had intrinsic mathematical interest. However, in 1977, a mathematically significant sentence of combinatorics, related to the so-called finite Ramsey theorem, was shown to be undecidable in S (see Kirby and Paris, 1977; Paris and Harrington, 1977; and Paris, 1978).

DEFINITION

A theory K is said to be *recursively axiomatizable* if there is a theory K* having the same theorems as K such that K* has a recursive axiom set.

COROLLARY 3.39

Let K be a theory in the language $\mathscr{L}_A$. If K is a consistent, recursively axiomatizable extension of RR, then K has an undecidable sentence.

Proof

Let K* be a theory having the same theorems as K and such that K* has a recursive axiom set. Conditions 1–5 of Proposition 3.38 hold for K*. Hence, a Rosser sentence for K* is undecidable in K* and, therefore, also undecidable in K.

An *effectively decidable* set of objects is a set for which there is a mechanical procedure that determines, for any given object, whether or not that object belongs to the set. By a *mechanical procedure* we mean a procedure that is carried out automatically without any need for originality or ingenuity in its application. On the other hand, a set A of natural numbers is said to be *recursive* if the property $x \in A$ is recursive.† The reader should be convinced after Chapter 5 that *the precise notion of recursive set corresponds to the intuitive idea of an effectively decidable set of natural numbers.* This hypothesis is known as *Church's thesis.*

Remember that a theory is said to be *axiomatic* if the set of its axioms is effectively decidable. Clearly, the set of axioms is effectively decidable if and only if the set of Gödel numbers of axioms is effectively decidable (since we can pass effectively from a wf to its Gödel number and, conversely, from the Gödel number to the wf). Hence, if we accept Church's thesis, to say that K has a recursive axiom set is equivalent to saying that K is an axiomatic theory, and, therefore, Corollary 3.39 shows RR is *essentially incomplete*, that is, that every consistent axiomatic extension of RR has an undecidable sentence. This result is very disturbing; it tells us that there is no complete axiomatization of arithmetic, that is, there is no way to set up an axiom system on the basis of which we can decide all problems of number theory.

Exercises

3.46 Church's thesis is usually taken in the form that *a number-theoretic function is effectively computable if and only if it is recursive.* Prove that this is equivalent to the form of Church's thesis given above.

†To say that $x \in A$ is recursive means that the characteristic function C_A is a recursive function, where $C_A(x) = 0$ if $x \in A$ and $C_A(x) = 1$ if $x \notin A$ (see page 180).

3.47 Let K be a true theory that satisfies the hypotheses of the Gödel–Rosser theorem. Determine whether a Rosser sentence $\mathscr{R}$ for K is true for the standard interpretation.

3.48 (Church, 1936b) Let Tr be the set of Gödel numbers of all wfs in the language $\mathscr{L}_A$ that are true for the standard interpretation. Prove that Tr is not recursive. (Hence, under the assumption of Church's thesis, there is no effective procedure for determining the truth or falsity of arbitrary sentences of arithmetic.)

3.49 Prove that there is no recursively axiomatizable theory that has Tr as the set of Gödel numbers of its theorems.

3.50 Let K be a theory with equality in the language $\mathscr{L}_A$ that satisfies conditions 4 and 5 on page 208. If every recursive relation is expressible in K, prove that every recursive function is representable in K.

Gödel's Second Theorem

Let K be an extension of S in the language $\mathscr{L}_A$ such that K has a recursive axiom set. Let $\mathscr{C}on_{\mathrm{K}}$ be the following closed wf of K:

$$(\forall x_1)(\forall x_2)(\forall x_3)(\forall x_4)\neg(\mathscr{P}f(x_1, x_3) \wedge \mathscr{P}f(x_2, x_4) \wedge \mathscr{N}eg(x_3, x_4))$$

For the standard interpretation, $\mathscr{C}on_{\mathrm{K}}$ asserts that there are no proofs in K of a wf and its negation, that is, that K is consistent.

Consider the following sentence:

$$(\mathbb{G}) \qquad \mathscr{C}on_{\mathrm{K}} \Rightarrow \mathscr{G}$$

where $\mathscr{G}$ is a Gödel sentence for K. Remember that $\mathscr{G}$ asserts that $\mathscr{G}$ is unprovable in K. Hence, ($\mathbb{G}$) states that, if K is consistent, then $\mathscr{G}$ is not provable in K. But that is just the first half of Gödel's incompleteness theorem. The metamathematical reasoning used in the proof of that theorem can be expressed and carried through within K itself, so that one obtains a proof in K of ($\mathbb{G}$) (see Hilbert & Bernays, 1939, pp. 285–328; Feferman, 1960). Thus, $\vdash_{\mathrm{K}} \mathscr{C}on_{\mathrm{K}} \Rightarrow \mathscr{G}$. But, by Gödel's incompleteness theorem, if K is consistent, $\mathscr{G}$ is not provable in K. Hence, *if* K *is consistent,* $\mathscr{C}on_{\mathrm{K}}$ *is not provable in* K.

This is *Gödel's second theorem* (1931). One can paraphrase it by stating that, if K is consistent, then the consistency of K cannot be proved within K, or, equivalently, a consistency proof of K must use ideas and methods that go beyond those available in K. Consistency proofs for S have been given by Gentzen (1936; 1938) and Schütte (1951), and these proofs do, in fact, employ notions and methods (for example, a portion of the theory of denumerable ordinal numbers) that apparently are not formalizable in S.

Gödel's second theorem is sometimes stated in the form that, if a 'sufficiently strong' theory K is consistent, then the consistency of K cannot be

proved within K. Aside from the vagueness of the 'sufficiently strong' (which can be made precise without much difficulty), the way in which the consistency of K is formulated is crucial. Feferman (1960, Cor. 5.10) has shown that there is a way of formalizing the consistency of S – say, $\mathscr{C}on^*_S$ – such that $\vdash_S \mathscr{C}on^*_S$. A precise formulation of Gödel second theorem may be found in Feferman (1960). (See Jeroslow (1971; 1972; 1973) for further clarification and development.)

In their proof of Gödel's second theorem, Hilbert and Bernays (1939) based their work on three so-called *derivability conditions*. For the sake of definiteness, we shall limit ourselves to the theory S, although everything we say also holds for recursively axiomatizable extensions of S. To formulate the Hilbert–Bernays results, let $\mathscr{B}ew(x_1)$ stand for $(\exists x_2)\mathscr{P}f(x_2, x_1)$. Thus, under the standard interpretation, $\mathscr{B}ew(x_1)$ means that there is a proof in S of the wf with Gödel number x_1; that is, the wf with Gödel number x_1 is provable in S.[†] Notice that a Gödel sentence $\mathscr{G}$ for S satisfies the fixed-point condition: $\vdash_S \mathscr{G} \Leftrightarrow \neg\mathscr{B}ew(\ulcorner\mathscr{G}\urcorner)$.

THE HILBERT-BERNAYS DERIVABILITY CONDITIONS[‡]

(HB1) If $\vdash_S \mathscr{C}$, then $\vdash_S \mathscr{B}ew(\ulcorner\mathscr{C}\urcorner)$.
(HB2) $\vdash_S \mathscr{B}ew(\ulcorner\mathscr{C} \Rightarrow \mathscr{D}\urcorner) \Rightarrow (\mathscr{B}ew(\ulcorner\mathscr{C}\urcorner) \Rightarrow \mathscr{B}ew(\ulcorner\mathscr{D}\urcorner))$
(HB3) $\vdash_S \mathscr{B}ew(\ulcorner\mathscr{C}\urcorner) \Rightarrow \mathscr{B}ew(\ulcorner\mathscr{B}ew(\ulcorner\mathscr{C}\urcorner)\urcorner)$

Here, $\mathscr{C}$ and $\mathscr{D}$ are arbitrary closed wfs of S. (HB1) is straightforward and (HB2) is an easy consequence of properties of $\mathscr{P}f$. However, (HB3) requires a careful and difficult proof. (A clear treatment may also be found in Boolos (1993, chap. 2), and in Shoenfield (1967, pp. 211–213).)

A Gödel sentence $\mathscr{G}$ for S asserts its own unprovability in S: $\vdash_S \mathscr{G} \Leftrightarrow \neg\mathscr{B}ew(\ulcorner\mathscr{G}\urcorner)$. We also can apply the fixed-point theorem to obtain a sentence $\mathscr{H}$ such that $\vdash_S \mathscr{H} \Leftrightarrow \mathscr{B}ew(\ulcorner\mathscr{H}\urcorner)$. $\mathscr{H}$ is called a *Henkin sentence* for S. $\mathscr{H}$ asserts its own provability in S. On intuitive grounds, it is not clear whether $\mathscr{H}$ is true for the standard interpretation, nor is it easy to determine whether $\mathscr{H}$ is provable, disprovable or undecidable in S. The problem was solved by Löb (1955) on the basis of Proposition 3.40 below. First, however, let us introduce the following convenient abbreviation.

[†]'Bew' consists of the first three letters of the German word *beweisbar*, which means 'provable'.

[‡]These three conditions are simplifications by Löb (1955) of the original Hilbert–Bernays conditions.

NOTATION

Let $\Box\mathscr{C}$ stand for $\mathscr{B}ew(\ulcorner\mathscr{C}\urcorner)$, where $\mathscr{C}$ is any wf. Then the Hilbert–Bernays derivability conditions become:

(HB1) If $\vdash_S \mathscr{C}$, then $\vdash_S \Box\mathscr{C}$.
(HB2) $\vdash_S \Box(\mathscr{C} \Rightarrow \mathscr{D}) \Rightarrow (\Box\mathscr{C} \Rightarrow \Box\mathscr{D})$
(HB3) $\vdash_S \Box\mathscr{C} \Rightarrow \Box\Box\mathscr{C}$

The Gödel sentence $\mathscr{G}$ and the Henkin sentence $\mathscr{H}$ satisfy the equivalences $\vdash_S \mathscr{G} \Leftrightarrow \neg\Box\mathscr{G}$ and $\vdash_S \mathscr{H} \Leftrightarrow \Box\mathscr{H}$.

PROPOSITION 3.40 (LÖB'S THEOREM)

Let $\mathscr{C}$ be a sentence of S. If $\vdash_S \Box\mathscr{C} \Rightarrow \mathscr{C}$, then $\vdash_S \mathscr{C}$.

Proof

Apply the fixed-point theorem to the wf $\mathscr{B}ew(x_1) \Rightarrow \mathscr{C}$ to obtain a sentence $\mathscr{L}$ such that $\vdash_S \mathscr{L} \Leftrightarrow (\mathscr{B}ew(\ulcorner\mathscr{L}\urcorner) \Rightarrow \mathscr{C})$. Thus, $\vdash_S \mathscr{L} \Leftrightarrow (\Box\mathscr{L} \Rightarrow \mathscr{C})$. Then we have the following derivation of $\mathscr{C}$.

1. $\vdash_S \mathscr{L} \Leftrightarrow (\Box\mathscr{L} \Rightarrow \mathscr{C})$	Obtained above
2. $\vdash_S \mathscr{L} \Rightarrow (\Box\mathscr{L} \Rightarrow \mathscr{C})$	1, biconditional elimination
3. $\vdash_S \Box(\mathscr{L} \Rightarrow (\Box\mathscr{L} \Rightarrow \mathscr{C}))$	2, (HB1)
4. $\vdash_S \Box\mathscr{L} \Rightarrow \Box(\Box\mathscr{L} \Rightarrow \mathscr{C})$	3, (HB2), MP
5. $\vdash_S \Box(\Box\mathscr{L} \Rightarrow \mathscr{C}) \Rightarrow (\Box\Box\mathscr{L} \Rightarrow \Box\mathscr{C})$	(HB2)
6. $\vdash_S \Box\mathscr{L} \Rightarrow (\Box\Box\mathscr{L} \Rightarrow \Box\mathscr{C})$	4, 5 tautology
7. $\vdash_S \Box\mathscr{L} \Rightarrow \Box\Box\mathscr{L}$	(HB3)
8. $\vdash_S \Box\mathscr{L} \Rightarrow \Box\mathscr{C}$	6, 7, tautology
9. $\vdash_S \Box\mathscr{C} \Rightarrow \mathscr{C}$	Hypothesis of the theorem
10. $\vdash_S \Box\mathscr{L} \Rightarrow \mathscr{C}$	8, 9, tautology
11. $\vdash_S \mathscr{L}$	1, 10, biconditional elimination
12. $\vdash_S \Box\mathscr{L}$	11, (HB1)
13. $\vdash_S \mathscr{C}$	10, 12, MP

COROLLARY 3.41

Let $\mathscr{H}$ be a Henkin sentence for S. Then $\vdash_S \mathscr{H}$ and $\mathscr{H}$ is true for the standard interpretation.

Proof

$\vdash_S \mathscr{H} \Leftrightarrow \Box\mathscr{H}$. By biconditional elimination, $\vdash_S \Box\mathscr{H} \Rightarrow \mathscr{H}$. So, by Löb's theorem, $\vdash_S \mathscr{H}$. Since $\mathscr{H}$ asserts that $\mathscr{H}$ is provable in S, $\mathscr{H}$ is true

Löb's theorem also enables us to give a proof of Gödel's second theorem for S.

PROPOSITION 3.42 (GÖDEL'S SECOND THEOREM)

If S is consistent, then not-$\vdash_S \mathscr{C}on_S$.

Proof

Assume S consistent. Since $\vdash_S 0 \neq \bar{1}$, the consistency of S implies not-$\vdash_S 0 = \bar{1}$. By Löb's theorem, not-$\vdash_S \Box(0 = \bar{1}) \Rightarrow 0 = \bar{1}$. Hence, by the tautology $\neg A \Rightarrow (A \Rightarrow B)$, we have:

$$(*) \qquad \text{not-} \vdash_S \neg\Box(0 = \bar{1})$$

But, since $\vdash_S 0 \neq \bar{1}$, (HB1) yields $\vdash_S \Box(0 \neq \bar{1})$. Then it is easy to show that $\vdash_S \mathscr{C}on_S \Rightarrow \neg\Box(0 = \bar{1})$. So, by $(*)$, not-$\vdash_S \mathscr{C}on_S$.

Boolos (1993) gives an elegant and extensive study of the fixed-point theorem and Löb's theorem in the context of an axiomatic treatment of provability predicates. Such an axiomatic approach was first proposed and developed by Magari (1975).

Exercises

3.51 Prove (HB1) and (HB2).

3.52 Give the details of the proof of $\vdash_S \mathscr{C}on_S \Rightarrow \neg\mathscr{B}ew(\ulcorner 0 = 1 \urcorner)$, which was used in the proof of Proposition 3.42.

3.53 If $\mathscr{G}$ is a Gödel sentence of S, prove $\vdash_S \mathscr{G} \Leftrightarrow \neg\mathscr{B}ew(\ulcorner 0 = \bar{1} \urcorner)$. (Hence, any two Gödel sentences for S are provably equivalent. This is an instance of a more general phenomenon of equivalence of fixed-point sentences, first noticed and verified independently by Bernardi (1975; 1976), De Jongh, and Sambin (1976). See Smoryński (1979; 1982).)

3.54 In each of the following cases, apply the fixed-point theorem for S to obtain a sentence of the indicated kind; determine whether that sentence is provable in S, disprovable in S, or undecidable in S; and determine the truth or falsity of the sentence for the standard interpretation.

(a) A sentence $\mathscr{C}$ that asserts its own decidability in S (that is, that $\vdash_S \mathscr{C}$ or $\vdash_S \neg\mathscr{C}$).

(b) A sentence that asserts its own undecidability in S.

(c) A sentence $\mathscr{C}$ asserting that not-$\vdash_S \neg\mathscr{C}$.

(d) A sentence $\mathscr{C}$ asserting that $\vdash_S \neg\mathscr{C}$.

3.6 RECURSIVE UNDECIDABILITY. CHURCH'S THEOREM

If K is a theory, let T_K be the set of Gödel numbers of theorems of K.

DEFINITIONS

K is said to be *recursively decidable* if T_K is a recursive set (that is, the property $x \in T_K$ is recursive). K is said to be *recursively undecidable* if T_K is not recursive. K is said to be *essentially recursively undecidable* if K and all consistent extensions of K are recursively undecidable.

If we accept Church's thesis, then recursive undecidability is equivalent to effective undecidability, that is, non-existence of a mechanical decision procedure for theoremhood. The non-existence of such a mechanical procedure means that ingenuity is required for determining whether arbitrary wfs are theorems.

Exercise

3.55 Prove that an inconsistent theory having a recursive vocabulary is recursively decidable.

PROPOSITION 3.43

Let K be a consistent theory with equality in the language $\mathscr{L}_A$ in which the diagonal function D is representable. Then the property $x \in T_K$ is not expressible in K.

Proof

Assume $x \in T_K$ is expressible in K by a wf $\mathscr{T}(x_1)$. Thus:

(a) If $n \in T_K$, $\vdash_K \mathscr{T}(\bar{n})$.
(b) If $n \notin T_K$, $\vdash_K \neg\mathscr{T}(\bar{n})$.

By the diagonalization lemma applied to $\neg\mathscr{T}(x_1)$, there is a sentence $\mathscr{C}$ such that $\vdash_K \mathscr{C} \Leftrightarrow \neg\mathscr{T}(\ulcorner\mathscr{C}\urcorner)$. Let q be the Gödel number of $\mathscr{C}$. So:

(c) $\vdash_K \mathscr{C} \Leftrightarrow \neg\mathscr{T}(\bar{q})$.

Case 1: $\vdash_K \mathscr{C}$. Then $q \in T_K$. By (a), $\vdash_K \mathscr{T}(\bar{q})$. But, from $\vdash_K \mathscr{C}$ and (c), by biconditional elimination, $\vdash_K \neg\mathscr{T}(\bar{q})$. Hence K is inconsistent, contradicting our hypothesis.

Case 2: not-$\vdash_K \mathscr{C}$. So, $q \notin T_K$. By (b), $\vdash_K \neg\mathscr{T}(\bar{q})$. Hence, by (c) and biconditional elimination, $\vdash_K \mathscr{C}$.

Thus, in either case a contradiction is reached.

DEFINITION

A set B of natural numbers is said to be *arithmetical* if there is a wf $\mathscr{B}(x)$ in the language $\mathscr{L}_A$, with one free variable x, such that, for every natural number n, $n \in B$ if and only if $\mathscr{B}(\bar{n})$ is true for the standard interpretation.

COROLLARY 3.44 [TARSKI'S THEOREM (1936)]

Let Tr be the set of Gödel numbers of wfs of S that are true for the standard interpretation. Then Tr is not arithmetical.

Proof

Let $\mathscr{N}$ be the extension of S that has as proper axioms all those wfs that are true for the standard interpretation. Since every theorem of $\mathscr{N}$ must be true for the standard interpretation, the theorems of $\mathscr{N}$ are identical with the axioms of $\mathscr{N}$. Hence, $T_{\mathscr{N}} = Tr$. Thus, for any closed wf $\mathscr{B}$, $\mathscr{B}$ holds for the standard interpretation if and only if $\vdash_{\mathscr{N}} \mathscr{B}$. It follows that a set B is arithmetical if and only if the property $x \in B$ is expressible in $\mathscr{N}$. We may assume that $\mathscr{N}$ is consistent because it has the standard interpretation as a model. Since every recursive function is representable in S, every recursive function is representable in $\mathscr{N}$ and, therefore, D is representable in $\mathscr{N}$. By Proposition 3.43, $x \in Tr$ is not expressible in $\mathscr{N}$. Hence, Tr is not arithmetical. (This result can be roughly paraphrased by saying that the notion of arithmetical truth is not arithmetically definable.)

PROPOSITION 3.45

Let K be a consistent theory with equality in the language $\mathscr{L}_A$ in which all recursive functions are representable. Assume also that $\vdash_K 0 \neq \bar{1}$. Then K is recursively undecidable.

Proof

D is primitive recursive and, therefore, representable in K. By Proposition 3.43, the property $x \in T_K$ is not expressible in K. By Proposition 3.13, the

characteristic function C_{T_K} is not representable in K. Hence, C_{T_K} is not a recursive function. Therefore, T_K is not a recursive set and so, by definition, K is recursively undecidable.

COROLLARY 3.46

RR is essentially recursively undecidable.

Proof

RR and all consistent extensions of RR satisfy the conditions on K in Proposition 3.45 and, therefore, are recursively undecidable. (We take for granted that RR is consistent because it has the standard interpretation as a model. More constructive consistency proofs can be given along the same lines as the proofs by Beth (1959, § 84) or Kleene (1952, § 79).)

We shall now show how this result can be used to give another derivation of the Gödel-Rosser theorem.

PROPOSITION 3.47

Let K be a theory with a recursive vocabulary. If K is recursively axiomatizable and recursively undecidable, then K is incomplete (i.e., K has an undecidable sentence).

Proof

By the recursive axiomatizability of K, there is a theory J with a recursive axiom set that has the same theorems as K. Since K and J have the same theorems, $T_K = T_J$ and, therefore, J is recursively undecidable, and K is incomplete if and only if J is incomplete. So, it suffices to prove J incomplete. Notice that, since K and J have the same theorems, J and K must have the same individual constants, function letters, and predicate letters (because all such symbols occur in logical axioms). Thus, the hypotheses of Propositions 3.26 an 3.28 hold for J. Moreover, J is consistent, since an inconsistent theory with a recursive vocabulary is recursively decidable.

Assume J is complete. Remember that, if x is the Gödel number of a wf, $\mathrm{Clos}(x)$ is the Gödel number of the closure of that wf. By Proposition 3.26 (16), Clos is a recursive function. Define:

$$\mathrm{H}(x) = \mu y[(\mathrm{Fml}(x) \wedge (\mathrm{Pf}(y, \mathrm{Clos}(x)) \vee \mathrm{Pf}(y, \mathrm{Neg}(\mathrm{Clos}(x))))) \vee \neg\mathrm{Fml}(x)]$$

Notice that, if x is not the Gödel number of a wf, $\mathrm{H}(x) = 0$. If x is the Gödel number of a wf $\mathscr{B}$, the closure of $\mathscr{B}$ is a closed wf and, by the completeness of J, there is a proof in J of either the closure of $\mathscr{B}$ or its negation. Hence,

H(x) is obtained by a legitimate application of the restricted μ-operator and, therefore, H is a recursive function. Recall that a wf is provable if and only if its closure is provable. So, $x \in T_J$ if and only if Pf(H(x), Clos(x)). But Pf(H(x), Clos(x)) is recursive. Thus, T_J is recursive, contradicting the recursive undecidability of J.

The intuitive idea behind this proof is the following. Given any wf $\mathscr{B}$, we form its closure $\mathscr{C}$ and start listing all the theorems in J. (Since PrAx is recursive, Church's thesis tells us that J is an axiomatic theory and, therefore, by the argument on page 86, we have an effective procedure for generating all the theorems.) If J is complete, either $\mathscr{C}$ or $\neg\mathscr{C}$ will eventually appear in the list of theorems. If $\mathscr{C}$ appears, $\mathscr{B}$ is a theorem. If $\neg\mathscr{C}$ appears, then, by the consistency of J, $\mathscr{C}$ will not appear among the theorems and, therefore, $\mathscr{B}$ is not a theorem. Thus, we have a decision procedure for theoremhood and, again by Church's thesis, J would be recursively decidable.

COROLLARY 3.48 (GÖDEL–ROSSER THEOREM)

Any consistent recursively axiomatizable extension of RR has undecidable sentences.

Proof

This is an immediate consequence of Corollary 3.46 and Proposition 3.47.

Exercises

3.56 Prove that a recursively decidable theory must be recursively axiomatizable.

3.57 Let K be any recursively axiomatizable true theory with equality. (So, $T_K \subseteq Tr$.) Prove that K has an undecidable sentence. [*Hint:* Use Proposition 3.47 and Exercise 3.48.]

3.58 Two sets A and B of natural numbers are said to be *recursively inseparable* if there is no recursive set C such that $A \subseteq C$ and $B \subseteq \bar{C}$. ($\bar{C}$ is the complement $\omega - C$.) Let K be any consistent theory with equality in the language $\mathscr{L}_A$ in which all recursive functions are representable and such that $\vdash_K 0 \neq \bar{1}$. Let Ref_K be the set of Gödel numbers of refutable wfs of K, that is, $\{x|\text{Neg}(x) \in T_K\}$. Prove that T_K and Ref_K are recursively inseparable.

DEFINITIONS

Let K_1 and K_2 be two theories in the same language.

(a) K_2 is called a *finite extension* of K_1 if and only if there is a set A of wfs and a finite set B of wfs such that: (1) the theorems of K_1 are precisely

the wfs derivable from A; and (2) the theorems of K_2 are precisely the wfs derivable from $A \cup B$.

(b) Let $K_1 \cup K_2$ denote the theory whose set of axioms is the union of the set of axioms of K_1 and the set of axioms of K_2. We say that K_1 and K_2 are *compatible* if $K_1 \cup K_2$ is consistent.

PROPOSITION 3.49

Let K_1 and K_2 be two theories in the same language. If K_2 is a finite extension of K_1 and if K_2 is recursively undecidable, then K_1 is recursively undecidable.

Proof

Let A be a set of axioms of K_1 and $A \cup \{\mathscr{B}_1, \ldots, \mathscr{B}_n\}$ a set of axioms for K_2. We may assume that $\mathscr{B}_1, \ldots, \mathscr{B}_n$ are closed wfs. Then, by Corollary 2.7, it is easy to see that a wf $\mathscr{C}$ is provable in K_2 if and only if $(\mathscr{B}_1 \wedge \ldots \wedge \mathscr{B}_n) \Rightarrow \mathscr{C}$ is provable in K_1. Let c be a Gödel number of $(\mathscr{B}_1 \wedge \ldots \wedge \mathscr{B}_n)$. Then b is a Gödel number of a theorem of K_2 when and only when $2^3 * c * 2^{11} * b * 2^5$ is a Gödel number of a theorem of K_1; that is, b is in T_{K_2} if and only if $2^3 * c * 2^{11} * b * 2^5$ is in T_{K_1}. Hence, if T_{K_1} were recursive, T_{K_2} would also be recursive, contradicting the recursive undecidability of K_2.

PROPOSITION 3.50

Let K be a theory in the language $\mathscr{L}_A$. If K is compatible with RR, then K is recursively undecidable.

Proof

Since K is comptatible with RR, the theory K∪RR is a consistent extension of RR. Therefore, by Corollary 3.46, K∪RR is recursively undecidable. Since RR has a finite number of axioms, K∪RR is a finite extension of K. Hence, by Proposition 3.49, K is recursively undecidable.

COROLLARY 3.51

Every true theory K is recursively undecidable.

Proof

K ∪ RR has the standard interpretation as a model and is, therefore, consistent. Thus, K is compatible with RR. Now apply Proposition 3.50.

COROLLARY 3.52

Let P_S be the predicate calculus in the language $\mathscr{L}_A$. Then P_S is recursively undecidable.

Proof

$P_S \cup RR = RR$. Hence, P_S is compatible with RR and, therefore, by Proposition 3.50, recursively undecidable.

By PF we mean the *full* first-order predicate calculus containing all predicate letters, function letters and individual constants. Let PP be the *pure* first-order predicate calculus, containing all predicate letters but no function letters or individual constants.

LEMMA 3.53

There is a recursive function h such that, for any wf $\mathscr{B}$ of PF having Gödel number u, there is a wf $\mathscr{B}'$ of PP having Gödel number $h(u)$ such that $\mathscr{B}$ is provable in PF if and only if $\mathscr{B}'$ is provable in PP.

Proof

Let $\mathscr{B}$ be a wf of PF. With the distinct function letters f_j^n in $\mathscr{B}$, associate distinct predicate letters A_r^{n+1} not occurring in $\mathscr{B}$, and with the distinct individual constants a_j in $\mathscr{B}$, associate distinct predicate letters A_k^1 not occurring in $\mathscr{B}$. Find the first individual constant a_j in $\mathscr{B}$ (if any). Let z be the first variable not in $\mathscr{B}$ and let $\mathscr{B}^*$ result from $\mathscr{B}$ by replacing all occurrences of a_j by z. Form the wf $\mathscr{B}_1$: $(\exists z)A_k^1(z) \Rightarrow (\exists z)(A_k^1(z) \wedge \mathscr{B}^*)$, where A_k^1 is the predicate letter associated with a_j. It is easy to check (see the proof of Proposition 2.28) that $\mathscr{B}$ is logically valid if and only if $\mathscr{B}_1$ is logically valid. Keep on performing similar transformations until a wf $\mathscr{C}$ without individual constants is reached; then $\mathscr{C}$ is logically valid if and only if $\mathscr{B}$ is logically valid. Next, take the leftmost term $f_\ell^n(t_1, \ldots, t_n)$ in $\mathscr{C}$, where $t_1, \ldots, t_n$ do not contain function letters. Let w be the first variable not in $\mathscr{C}$, let $\mathscr{C}^\#$ result from $\mathscr{C}$ by replacing $f_\ell^n(t_1, \ldots, t_n)$ by w, and let $\mathscr{C}_1$ be the wf $(\exists w)A_r^{n+1}(w, t_1, \ldots, t_n) \Rightarrow (\exists w)(A_r^{n+1}(w, t_1, \ldots, t_n) \wedge \mathscr{C}^\#)$, where A_r^{n+1} is the predicate letter associated with f_ℓ^n. It is easy to verify that $\mathscr{C}$ is logically valid if and only if $\mathscr{C}_1$ is logically valid. Repeat the same transformation on $\mathscr{C}_1$, and so on, until a wf $\mathscr{B}'$ is reached that contains no function letters. Then $\mathscr{B}'$ is a wf of PP, and $\mathscr{B}'$ is logically valid if and only if $\mathscr{B}$ is logically valid. By Gödel's completeness theorem (Corollary 2.19), $\mathscr{B}$ is logically valid if and only if $\vdash_{PF} \mathscr{B}$, and $\mathscr{B}'$ is logically valid if and only if $\vdash_{PP} \mathscr{B}'$. Now, if u is the

Gödel number of $\mathscr{B}$, let $h(u)$ be the Gödel number of $\mathscr{B}'$. When u is not the Gödel number of a wf of PF, define $h(u)$ to be 0. Clearly, h is effectively computable because we have described an effective procedure for obtaining $\mathscr{B}'$ from $\mathscr{B}$. Therefore, by Church's thesis, h is recursive. Alternatively, an extremely diligent reader could avoid the use of Church's thesis by 'arithmetizing' all the steps described above in the computation of h.

PROPOSITION 3.54 (CHURCH'S THEOREM (1936a))

PF and PP are recursively undecidable.

Proof

(a) By Gödel's completeness theorem, a wf $\mathscr{B}$ of $\mathrm{P_S}$ is provable in $\mathrm{P_S}$ if and only if $\mathscr{B}$ is logically valid, and $\mathscr{B}$ is provable in PF if and only if $\mathscr{B}$ is logically valid. Hence, $\vdash_{\mathrm{P_S}} \mathscr{B}$ if and only if $\vdash_{\mathrm{PF}} \mathscr{B}$. However, the set $\mathrm{Fml_{P_S}}$ of Gödel numbers of wfs of $\mathrm{P_S}$ is recursive. Then $T_{\mathrm{P_S}} = T_{\mathrm{PF}} \cap \mathrm{Fml_{P_S}}$, where $T_{\mathrm{P_S}}$ and T_{PF} are, respectively, the sets of Gödel numbers of the theorems of P_s and PF. If T_{PF} were recursive, $T_{\mathrm{P_S}}$ would be recursive, contradicting Corollary 3.52. Therefore, PF is recursively undecidable.

(b) By Lemma 3.53, u is in T_{PF} if and only if $h(u)$ is in $\mathrm{T_{PP}}$. Since h is recursive, the recursiveness of T_{PP} would imply the recursiveness of $\mathrm{T_{PF}}$, contradicting (a). Thus, T_{PP} is not recursive; that is, PP is recursively undecidable.

If we accept Church's thesis, then 'recursively undecidable' can be replaced everywhere by 'effectively undecidable'. In particular, Proposition 3.54 states that there is no decision procedure for recognizing theoremhood, either for the pure predicate calculus PP or the full predicate calculus PF. By Gödel's completeness theorem, this implies that *there is no effective method for determining whether any given wf is logically valid.*

Exercises

3.59$^{\mathrm{D}}$

(a) By a wf of the pure monadic predicate calculus (PMP) we mean a wf of the pure predicate calculus that does not contain predicate letters of more than one argument. Show that, in contrast to Church's theorem, there is an effective procedure for determining whether a wf of PMP is logically valid. [*Hint*: Let $B_1, B_2, \ldots, B_k$ be the distinct predicate letters in a wf $\mathscr{B}$. Then $\mathscr{B}$ is logically valid if and only if $\mathscr{B}$ is true for every interpretation with at most 2^k elements. (In fact, assume $\mathscr{B}$ is true for

every interpretation with at most 2^k elements, and let M be any interpretation. For any elements b and c of the domain D of M, call b and c *equivalent* if the truth values of $B_1(b), B_2(b), \ldots, B_k(b)$ in M are, respectively, the same as those of $B_1(c), B_2(c), \ldots, B_k(c)$. This defines an equivalence relation in D, and the corresponding set of equivalence classes has at most 2^k members and can be made the domain of an interpretation M^* by defining interpretations of $B_1, \ldots, B_k$, in the obvious way, on the equivalence classes. By induction on the length of wfs $\mathscr{C}$ that contain no predicate letters other than $B_1, \ldots, B_k$, one can show that $\mathscr{C}$ is true for M if and only if it is true for M^*. Since $\mathscr{B}$ is true for M^*, it is also true for M. Hence, $\mathscr{B}$ is true for every interpretation.) Note also that whether $\mathscr{B}$ is true for every interpretation that has at most 2^k elements can be effectively determined.]†

(b) Prove that a wf $\mathscr{B}$ of PMP is logically valid if and only if $\mathscr{B}$ is true for all finite interpretations. (This contrasts with the situation in the pure predicate calculus; see Exercise 2.56 on page 93.)

3.60 If a theory K^* is consistent, if every theorem of an essentially recursively undecidable theory K_1 is a theorem of K^*, and if the property $\mathrm{Fml}_{K_1}(y)$ is recursive, prove that K^* is essentially recursively undecidable.

3.61 (Tarski, Mostowski and Robinson, 1953, I)

(a) Let K be a theory with equality. If a predicate letter A_j^n, a function letter f_j^n and an individual constant a_j are not symbols of K, then by *possible definitions* of A_j^n, f_j^n and a_j in K we mean, respectively, expressions of the form

(i) $(\forall x_1) \ldots (\forall x_n)(A_j^n(x_1, \ldots, x_n) \Leftrightarrow \mathscr{B}(x_1, \ldots, x_n))$

(ii) $(\forall x_1) \ldots (\forall x_n)(\forall y)(f_j^n(x_1, \ldots, x_n) = y \Leftrightarrow \mathscr{C}(x_1, \ldots, x_n, y))$

(iii) $(\forall y)(a_j = y \Leftrightarrow \mathscr{D}(y))$

where $\mathscr{B}$, $\mathscr{C}$ and $\mathscr{D}$ are wfs of K; moreover, in case (ii), we must also have $\vdash_K (\forall x_1) \ldots (\forall x_n)(\exists_1 y)\mathscr{C}(x_1, \ldots, x_n, y)$, and, in case (iii), $\vdash_K (\exists_1 y)\mathscr{D}(y)$. If K is consistent, prove that addition of any possible definitions to K as new axioms (using only one possible definition for each symbol) yields a consistent theory K′, and K′ is recursively undecidable if and only if K is.

(b) By a *non-logical constant* we mean a predicate letter, function letter or individual constant. Let K_1 be a theory with equality that has a finite number of non-logical constants. Then K_1 is said to be *interpretable* in a theory with equality K if we can associate with each non-logical constant of K_1 that is not a non-logical constant of K a possible definition

†The result in this exercise is, in a sense, the best possible. By a theorem of Kalmár (1936), there is an effective procedure producing for each wf $\mathscr{B}$ of the pure predicate calculus another wf $\mathscr{B}_2$ of the pure predicate calculus such that $\mathscr{B}_2$ contains only one predicate letter, a binary one, and such that $\mathscr{B}$ is logically valid if and only if $\mathscr{B}_2$ is logically valid. (For another proof, see Church, 1956, § 47.) Hence, by Church's theorem, there is no decision procedure for logical validity of wfs that contain only binary predicate letters. (For another proof, see Exercise 4.68 on page 271.)

in K such that, if K* is the theory obtained from K by adding ... possible definitions as axioms, then every axiom (and hence every orem) of K_1 is a theorem of K*. Notice that, if K_1 is interpretable ... it is interpretable in every extension of K. Prove that, if K_1 is in... pretable in K and K is consistent, and if K_1 is essentially recurs... undecidable, then K is essentially recursively undecidable.

3.62 Let K be a theory with equality and A_j^1 a monadic predicate lette... in K. Given a closed wf $\mathscr{C}$, let $\mathscr{C}^{(A_j^1)}$ (called the *relativization* of $\mathscr{C}$ with re... to A_j^1) be the wf obtained from $\mathscr{C}$ by replacing every subformula (sta... from the smallest subformulas) of the form $(\forall x)\mathscr{B}(x)$ by $(\forall x)(A_j^1(x) \Rightarrow \mathscr{B}$... Let the proper axioms of a new theory with equality $K^{A_j^1}$ be: (i) all wfs $\mathscr{C}$... where $\mathscr{C}$ is the closure of any proper axiom of K; (ii) $(\exists x)A_j^1(x)$; (iii) A_j^1... for each individual constant a_m of K; and (iv) $A_j^1(x_1) \wedge$... $A_j^1(x_n) \Rightarrow A_j^1(f_k^n(x_1, \ldots, x_n))$ for any function letter f_k^n of K. Prove the f... lowing.

(a) As proper axioms of $K^{A_j^1}$ we could have taken all wfs $\mathscr{C}^{(A_j^1)}$, where $\mathscr{C}$... the closure of any theorem of K.

(b) $K^{A_j^1}$ is interpretable in K.

(c) $K^{A_j^1}$ is consistent if and only if K is consistent.

(d) $K^{A_j^1}$ is essentially recursively undecidable if and only if K is (Tars... Mostowski and Robinson, 1953, pp. 27–28).

3.63 K is said to be *relatively interpretable* in K′ if there is some predic... letter A_j^1 not in K such that $K^{A_j^1}$ is interpretable in K′. If K is relativ... interpretable in a consistent theory K′ and K is essentially recursively u... decidable, prove that K′ is essentially recursively undecidable.

3.64 Call a theory K in which RR is relatively interpretable *sufficien... strong*. Prove that any sufficiently strong consistent theory K is essentia... recursively undecidable, and, if K is also recursively axiomatizable, pro... that K is incomplete. Roughly speaking, we may say that K is sufficien... strong if the notions of natural number, 0, 1, addition and multiplication a... 'definable' in K in such a way that the axioms of RR (relativized to t... 'natural numbers' of K) are provable in K. Clearly, any theory adequate ... present-day mathematics will be sufficiently strong and so, if it is consistent... then it will be recursively undecidable and, if it is recursively axiomatizable... then it will be incomplete. If we accept Church's thesis, this implies that a... consistent sufficiently strong theory will be effectively undecidable and, if ... is axiomatic, it will have undecidable sentences. (Similar results also hold f... higher-order theories; for example, see Gödel, 1931.) *This destroys all h... for a consistent and complete axiomatization of mathematics.*

Axiomatic Set Theory

4

1 AN AXIOM SYSTEM

prime reason for the increase in importance of mathematical logic in the ventieth century was the discovery of the paradoxes of set theory and the :d for a revision of intuitive (and contradictory) set theory. Many dif- rent axiomatic theories have been proposed to serve as a foundation for t theory but, no matter how they may differ at the fringes, they all have as common core the fundamental theorems that mathematicians require for ;eir daily work. We make no claim about the superiority of the system we ;all use except that, from a notational and conceptual standpoint, it is a onvenient basis for present-day mathematics.

We shall describe a first-order theory NBG, which is basically a system of he same type as one originally proposed by von Neumann (1925; 1928) and later thoroughly revised and simplified by R. Robinson (1937), Bernays 1937–1954); and Gödel (1940) (We shall follow Godel's monograph to a reat extent, although there will be some significant differences.)

NBG has a single predicate letter A_2^2 but no function letter, or individual constants.† In order to conform to the notation in Bernays (1937–1954) and Gödel (1940), we shall use capital italic letters X_1, X_2, X_3, ... as variables nstead of x_1, x_2, x_3, ... (As usual, we shall use X, Y, Z, ... to represent rbitrary variables.) We shall abbreviate $A_2^2(X, Y)$ by $X \in Y$, and $\neg A_2^2(X, Y)$ y $X \notin Y$.

Intuitively, $\in$ is to be thought of as the membership relation and the values of the variables are to be thought of as classes. Classes are certain collections of objects. Some properties determine classes, in the sense that a property P may determine a class of all those objects that possess that property. This 'interpretation' is as imprecise as the notions of 'collection' nd 'property'. The axioms will reveal more about what we have in mind. hey will provide us with the classes we need in mathematics and appear nodest enough so that contradictions are not derivable from them.

†We use A_2^2 instead of A_1^2 because the latter was used previously for the equality elation.

Let us define equality in the following way.

DEFINITION

$$X = Y \text{ for } (\forall Z)(Z \in X \Leftrightarrow Z \in Y)^{\dagger}$$

Thus, two classes are equal when and only when they have the same members.

DEFINITIONS

$X \subseteq Y$	for	$(\forall Z)(Z \in X \Rightarrow Z \in Y)$	(inclusion)
$X \subset Y$	for	$X \subseteq Y \wedge X \neq Y$	(proper inclusion)

When $X \subseteq Y$, we say that X is a *subclass* of Y. When $X \subset Y$, we say that X is a *proper subclass* of Y.

As easy consequences of these definitions, we have the following.

PROPOSITION 4.1[‡]

(a) $\vdash X = Y \Leftrightarrow (X \subseteq Y \wedge Y \subseteq X)$
(b) $\vdash X = X$
(c) $\vdash X = Y \Rightarrow Y = X$
(d) $\vdash X = Y \Rightarrow (Y = Z \Rightarrow X = Z)$

We shall now present the proper axioms of NBG, interspersing among the axioms some additional definitions and various consequences of the axioms.

We shall define a class to be a *set* if it is a member of some class. Those classes that are not sets are called *proper classes*.

DEFINITIONS

$\mathrm{M}(X)$	for	$(\exists Y)(X \in Y)$	(X is a set)
$\mathrm{Pr}(X)$	for	$\neg \mathrm{M}(X)$	(X is a proper class)

It will be seen later that the usual derivations of the paradoxes now no longer lead to contradictions but only yield the results that various classes are proper classes, not sets. The sets are intended to be those safe, comfortable classes that are used by mathematicians in their daily work, whereas

[†]As usual, Z is to be the first variable different from X and Y.
[‡]The subscript NBG will be omitted from $\vdash_{\mathrm{NBG}}$ in the rest of this chapter.

proper classes are thought of as monstrously large collections that, if permitted to be sets (i.e., allowed to belong to other classes), would engender contradictions.

Exercise 4.1 Prove $\vdash X \in Y \Rightarrow \mathrm{M}(X)$.

The system NBG is designed to handle classes, not concrete individuals.[†] The reason for this is that mathematics has no need for objects such as cows and molecules; all mathematical objects and relations can be formulated in terms of classes alone. If non-classes are required for applications to other sciences, then the system NBG can be modified slightly so as to apply to both classes and non-classes alike (see the system UR in Section 4.6 below).

Let us introduce lower-case letters $x_1, x_2, \ldots$ as special restricted variables for sets. In other words, $(\forall x_j)\mathscr{B}(x_j)$ stands for $(\forall X)(\mathrm{M}(X) \Rightarrow \mathscr{B}(X))$, that is, $\mathscr{B}$ holds for all sets, and $(\exists x_j)\mathscr{B}(x_j)$ stands for $(\exists X)(\mathrm{M}(X) \wedge \mathscr{B}(X))$, that is, $\mathscr{B}$ holds for some set. As usual, the variable X used in these definitions should be the first one that does not occur in $\mathscr{B}(x_j)$. We shall use $x, y, z, \ldots$ to stand for arbitrary set variables.

Example

$$(\forall X_1)(\forall x)(\exists y)(\exists X_3)(X_1 \in x \wedge y \in X_3) \text{ stands for}$$
$$(\forall X_1)(\forall X_2)(\mathrm{M}(X_2) \Rightarrow (\exists X_4)(\mathrm{M}(X_4) \wedge (\exists X_3)(X_1 \in X_2 \wedge X_4 \in X_3)))$$

Exercise 4.2

Prove that $\vdash X = Y \Leftrightarrow (\forall z)(z \in X \Leftrightarrow z \in Y)$. This is the so-called *extensionality principle:* two classes are equal when and only when they contain the same *sets* as members.

AXIOM T

$$X_1 = X_2 \Rightarrow (X_1 \in X_3 \Leftrightarrow X_2 \in X_3)$$

This axiom tells us that equal classes belong to the same classes.

Exercise

4.3 Prove that $\vdash \mathrm{M}(Z) \wedge Z = Y \Rightarrow \mathrm{M}(Y)$.

[†]If there were concrete individuals (that is, objects that are not classes), then the definition of equality would have to be changed, since all such individuals have the same members (namely, none at all).

PROPOSITION 4.2

NBG is a first-order theory with equality.

Proof

Use Proposition 4.1, axiom T, the definition of equality, and the discussion on page 99.

AXIOM P (PAIRING AXIOM)

$$(\forall x)(\forall y)(\exists z)(\forall u)(u \in z \Leftrightarrow u = x \vee u = y)$$

Thus, for any sets x and y, there is a set z that has x and y as its only members.

Exercises

4.4 Prove $\vdash (\forall x)(\forall y)(\exists_1 z)(\forall u)(u \in z \Leftrightarrow u = x \vee u = y)$. This asserts that there is a unique set z, called the *unordered pair* of x and y, such that z has x and y as its only members. Use axiom P and the extensionality principle.
4.5 Prove $\vdash (\forall X)(\mathrm{M}(X) \Leftrightarrow (\exists y)(X \in y))$.
4.6 Prove $\vdash (\exists X)\mathrm{Pr}(X) \Rightarrow \neg(\forall Y)(\forall Z)(\exists W)(\forall U)(U \in Z \Leftrightarrow U = X \vee U = Y)$.

AXIOM N (NULL SET)

$$(\exists x)(\forall y)(y \notin x)$$

Thus, there is a set that has no members. From axiom N and the extensionality principle, there is a unique set that has no members -- that is, $\vdash (\exists_1 x)(\forall y)(y \notin x)$. Therefore, we can introduce a new individual constant $\emptyset$ by means of the following condition.

DEFINITION

$$(\forall y)(y \notin \emptyset)$$

It then follows from axiom N and Exercise 4.3 that $\emptyset$ is a set.

Since we have (by Exercise 4.4) the uniqueness condition for the unordered pair, we can introduce a new function letter $g(x, y)$ to designate the unordered pair of x and y. In accordance with the traditional notation, we shall write $\{x, y\}$ instead of $g(x, y)$. Notice that we have to define a unique

value for $\{X, Y\}$ for any classes X and Y, not only for sets x and y. We shall let $\{X, Y\}$ be $\emptyset$ whenever X is not a set or Y is not a set. One can prove: $\vdash (\exists_1 Z)([(\neg \mathrm{M}(X) \vee \neg \mathrm{M}(Y)) \wedge Z = \emptyset] \vee [\mathrm{M}(X) \wedge \mathrm{M}(Y) \wedge (\forall u)(u \in Z \Leftrightarrow u = X \vee u = Y)])$. This justifies the introduction of a term $\{X, Y\}$ satisfying the following condition:

$$[\mathrm{M}(X) \wedge \mathrm{M}(Y) \wedge (\forall u)(u \in \{X, Y\} \Leftrightarrow u = X \vee u = Y)] \vee [(\neg \mathrm{M}(X) \vee \neg \mathrm{M}(Y)) \wedge \{X, Y\} = \emptyset]$$

One can then prove $\vdash (\forall x)(\forall y)(\forall u)(u \in \{x, y\} \Leftrightarrow u = x \vee u = y)$ and $\vdash (\forall X)(\forall Y)\mathrm{M}(\{X, Y\})$.

DEFINITION

$$\{X\} \text{ for } \{X, X\}$$

For a set x, $\{x\}$ is called the *singleton* of x. It is a set that has x as its only member.

In connection with these definitions, the reader should review Section 2.9 and, in particular, Proposition 2.28, which assures us that the introduction of new individual constants and function letters, such as $\emptyset$ and $\{X, Y\}$, adds nothing essentially new to the theory NBG.

Exercise

4.7 (a) Prove $\vdash \{X, Y\} = \{Y, X\}$.
(b) Prove $\vdash (\forall x)(\forall y)(\{x\} = \{y\} \Rightarrow x = y)$.

DEFINITION

$\langle X, Y \rangle$ for $\{\{X\}, \{X, Y\}\}$
For sets x and y, $\langle x, y \rangle$ is called the *ordered pair* of x and y.

The definition of $\langle X, Y \rangle$ does not have any intrinsic intuitive meaning. It is just a convenient way (discovered by Kuratowski, 1921) to define ordered pairs so that one can prove the characteristic property of ordered pairs expressed in the following proposition.

PROPOSITION 4.3

$$\vdash (\forall x)(\forall y)(\forall u)(\forall v)(\langle x, y \rangle = \langle u, v \rangle \Rightarrow x = u \wedge y = v)$$

Proof

Assume $\langle x,y\rangle = \langle u,v\rangle$. Then $\{\{x\},\{x,y\}\} = \{\{u\},\{u,v\}\}$. Since $\{x\} \in \{\{x\},\{x,y\}\}$, $\{x\} \in \{\{u\},\{u,v\}\}$. Hence, $\{x\} = \{u\}$ or $\{x\} = \{u,v\}$. In either case, $x = u$. Now, $\{u,v\} \in \{\{u\},\{u,v\}\}$; so, $\{u,v\} \in \{\{x\},\{x,y\}\}$. Then $\{u,v\} = \{x\}$ or $\{u,v\} = \{x,y\}$. Similarly, $\{x,y\} = \{u\}$ or $\{x,y\} = \{u,v\}$. If $\{u,v\} = \{x\}$ and $\{x,y\} = \{u\}$, then $x = y = u = v$; if not, $\{u,v\} = \{x,y\}$. Hence, $\{u,v\} = \{u,y\}$. So, if $v \neq u$, then $y = v$; if $v = u$, then $y = v$. Thus, in all cases, $y = v$.

Notice that the converse of Proposition 4.3 holds by virtue of the substitutivity of equality.

Exercise

4.8 (a) Show that, instead of the definition of an ordered pair given in the text, we could have used $\langle X,Y\rangle = \{\{\emptyset,X\},\{\{\emptyset\},Y\}\}$; that is, Proposition 4.3 would still be provable with this new meaning of $\langle X,Y\rangle$.

(b) Show that the ordered pair also could be defined as $\{\{\emptyset,\{X\}\},\{\{Y\}\}\}$. (This was the first such definition, discovered by Wiener (1914). For a thorough analysis of such definitions, see A. Oberschelp (1991).)

We now extend the definition of ordered pairs to ordered n-tuples.

DEFINITIONS

$$\langle X\rangle = X$$
$$\langle X_1,\ldots,X_n,X_{n+1}\rangle = \langle\langle X_1,\ldots,X_n\rangle,X_{n+1}\rangle$$

Thus, $\langle X,Y,Z\rangle = \langle\langle X,Y\rangle,Z\rangle$ and $\langle X,Y,Z,U\rangle = \langle\langle\langle X,Y\rangle,Z\rangle,U\rangle$.

It is easy to establish the following generalization of Proposition 4.3.

$$\vdash (\forall x_1)\ldots(\forall x_n)(\forall y_1)\ldots(\forall y_n)(\langle x_1,\ldots,x_n\rangle = \langle y_1,\ldots,y_n\rangle \Rightarrow x_1 = y_1 \wedge \ldots \wedge x_n = y_n)$$

AXIOMS OF CLASS EXISTENCE

(B1) $(\exists X)(\forall u)(\forall v)(\langle u,v\rangle \in X \Leftrightarrow u \in v)$ ($\in$-relation)
(B2) $(\forall X)(\forall Y)(\exists Z)(\forall u)(u \in Z \Leftrightarrow u \in X \wedge u \in Y)$ (intersection)
(B3) $(\forall X)(\exists Z)(\forall u)(u \in Z \Leftrightarrow u \notin X)$ (complement)
(B4) $(\forall X)(\exists Z)(\forall u)(u \in Z \Leftrightarrow (\exists v)(\langle u,v\rangle \in X))$ (domain)
(B5) $(\forall X)(\exists Z)(\forall u)(\forall v)(\langle u,v\rangle \in Z \Leftrightarrow u \in X)$
(B6) $(\forall X)(\exists Z)(\forall u)(\forall v)(\forall w)(\langle u,v,w\rangle \in Z \Leftrightarrow \langle v,w,u\rangle \in X)$
(B7) $(\forall X)(\exists Z)(\forall u)(\forall v)(\forall w)(\langle u,v,w\rangle \in Z \Leftrightarrow \langle u,w,v\rangle \in X)$

From axioms (B2)–(B4) and the extensionality principle, we obtain:

$\vdash (\forall X)(\forall Y)(\exists_1 Z)(\forall u)(u \in Z \Leftrightarrow u \in X \wedge u \in Y)$

$\vdash (\forall X)(\exists_1 Z)(\forall u)(u \in Z \Leftrightarrow u \notin X)$

$\vdash (\forall X)(\exists_1 Z)(\forall u)(u \in Z \Leftrightarrow (\exists v)(\langle u, v\rangle \in X))$

These results justify the introduction of new function letters: $\cap$, $^-$ and $\mathscr{D}$.

DEFINITIONS

$(\forall u)(u \in X \cap Y \Leftrightarrow u \in X \wedge u \in Y)$	(intersection of X and Y)
$(\forall u)(u \in \overline{X} \Leftrightarrow u \notin X)$	(complement of X)
$(\forall u)(u \in \mathscr{D}(X) \Leftrightarrow (\exists v)(\langle u, v\rangle \in X))$	(domain of X)
$X \cup Y = \overline{\overline{X} \cap \overline{Y}}$	(union of X and Y)
$V = \overline{\emptyset}$	(universal class)[†]
$X - Y = X \cap \overline{Y}$	(difference of X and Y)

Exercises

4.9 Prove:

(a) $\vdash (\forall u)(u \in X \cup Y \Leftrightarrow u \in X \vee u \in Y)$

(b) $\vdash (\forall u)(u \in V)$

(c) $\vdash (\forall u)(u \in X - Y \Leftrightarrow u \in X \wedge u \notin Y)$

4.10 Prove:

(a) $\vdash X \cap Y = Y \cap X$

(b) $\vdash X \cup Y = Y \cup X$

(c) $\vdash X \subseteq Y \Leftrightarrow X \cap Y = X$

(d) $\vdash X \subseteq Y \Leftrightarrow X \cup Y = Y$

(e) $\vdash (X \cap Y) \cap Z = X \cap (Y \cap Z)$

(f) $\vdash (X \cup Y) \cup Z = X \cup (Y \cup Z)$

(g) $\vdash X \cap X = X$

(h) $\vdash X \cup X = X$

(i) $\vdash X \cap \emptyset = \emptyset$

(j) $\vdash X \cup \emptyset = X$

(k) $\vdash X \cap V = X$

(l) $\vdash X \cup V = V$

(m) $\vdash \overline{X \cup Y} = \overline{X} \cap \overline{Y}$

(n) $\vdash \overline{X \cap Y} = \overline{X} \cup \overline{Y}$

(o) $\vdash X - X = \emptyset$

(p) $\vdash V - X = \overline{X}$

(q) $\vdash X - (X - Y) = X \cap Y$

(r) $\vdash Y \subseteq \overline{X} \Rightarrow X - Y = X$

(s) $\vdash \overline{\overline{X}} = X$

(t) $\vdash \overline{V} = \emptyset$

(u) $\vdash X \cap (Y \cup Z) = (X \cap Y) \cup (X \cap Z)$

(v) $\vdash X \cup (Y \cap Z) = (X \cup Y) \cap (X \cup Z)$

4.11 Prove the following wfs.

(a) $\vdash (\forall X)(\exists Z)(\forall u)(\forall v)(\langle u, v\rangle \in Z \Leftrightarrow \langle v, u\rangle \in X)$ [*Hint:* Apply axioms (B5),(B7),(B6) and (B4) successively.]

(b) $\vdash (\forall X)(\exists Z)(\forall u)(\forall v)(\forall w)(\langle u, v, w\rangle \in Z \Leftrightarrow \langle u, w\rangle \in X)$ [*Hint:* Use (B5) and (B7).]

(c) $\vdash (\forall X)(\exists Z)(\forall v)(\forall x_1) \ldots (\forall x_n)(\langle x_1, \ldots, x_n, v\rangle \in Z \Leftrightarrow \langle x_1, \ldots, x_n\rangle \in X)$ [*Hint:* Use (B5).]

[†]It will be shown later that V is a proper class, that is, V is not a set.

(d) $\vdash (\forall X)(\exists Z)(\forall v_1)\dots(\forall v_m)(\forall x_1)\dots(\forall x_n)(\langle x_1,\dots,x_n,v_1,\dots,v_m\rangle \in Z \Leftrightarrow \langle x_1,\dots,x_n\rangle \in X)$ [*Hint:* Iteration of part (c).]

(e) $\vdash (\forall X)(\exists Z)(\forall v_1)\dots(\forall v_m)(\forall x_1)\dots(\forall x_n)(\langle x_1,\dots,x_{n-1},v_1,\dots,v_m,x_n\rangle \in Z \Leftrightarrow \langle x_1,\dots,x_n\rangle \in X)$ [*Hint:* For $m=1$, use (b), substituting $\langle x_1,\dots,x_{n-1}\rangle$ for u and x_n for w; the general case then follows by iteration.]

(f) $\vdash (\forall X)(\exists Z)(\forall x)(\forall v_1)\dots(\forall v_m)(\langle v_1,\dots,v_m,x\rangle \in Z \Leftrightarrow x \in X)$ [*Hint:* Use (B5) and part (a).]

(g) $\vdash (\forall X)(\exists Z)(\forall x_1)\dots(\forall x_n)(\langle x_1,\dots,x_n\rangle \in Z \Leftrightarrow (\exists y)(\langle x_1,\dots,x_n,y\rangle \in X))$ [*Hint:* In (B4), substitute $\langle x_n,\dots,x_n\rangle$ for u and y for v.]

(h) $\vdash (\forall X)(\exists Z)(\forall u)(\forall v)(\forall w)(\langle v,u,w\rangle \in Z \Leftrightarrow \langle u,w\rangle \in X)$ [*Hint:* Substitute $\langle u,w\rangle$ for u in (B5) and apply (B6).]

(i) $\vdash (\forall X)(\exists Z)(\forall v_1)\dots(\forall v_k)(\forall u)(\forall w)(\langle v_1,\dots,v_k,u,w\rangle \in Z \Leftrightarrow \langle u,w\rangle \in X)$ [*Hint:* Substitute $\langle v_1,\dots,v_k\rangle$ for v in part (h).]

Now we can derive a general class existence theorem. By a *predicative wf* we mean a wf $\varphi(X_1,\dots,X_n,Y_1,\dots,Y_m)$ whose variables occur among $X_1,\dots,X_n,Y_1,\dots,Y_m$ and in which only set variables are quantified (i.e., φ can be abbreviated in such a way that only set variables are quantified).

Examples
$(\exists x_1)(x_1 \in Y_1)$ is predicative, whereas $(\exists Y_1)(x_1 \in Y_1)$ is not predicative.

PROPOSITION 4.4 (CLASS EXISTENCE THEOREM)

Let $\varphi(X_1,\dots,X_n,Y_1,\dots,Y_m)$ be a predicative wf. Then
$\vdash (\exists Z)(\forall x_1)\dots(\forall x_n)(\langle x_1,\dots,x_n\rangle \in Z \Leftrightarrow \varphi(x_1,\dots,x_n,Y_1,\dots,Y_m))$.

Proof

We shall consider only wfs φ in which no wf of the form $Y_i \in W$ occurs, since $Y_i \in W$ can be replaced by $(\exists x)(x = Y_i \wedge x \in W)$, which is equivalent to $(\exists x)[(\forall z)(z \in x \Leftrightarrow z \in Y_i) \wedge x \in W]$. Moreover, we may assume that φ contains no wf of the form $X \in X$, since this may be replaced by $(\exists u)(u = X \wedge u \in X)$, which is equivalent to $(\exists u)[(\forall z)(z \in u \Leftrightarrow z \in X) \wedge u \in X]$. We shall proceed now by induction on the number k of connectives and quantifiers in φ (written with restricted set variables).

Base: $k=0$. Then φ has the form $x_i \in x_j$ or $x_j \in x_i$ or $x_i \in Y_\ell$, where $1 \leqslant i < j \leqslant n$. For $x_i \in x_j$, axiom (B1) guarantees that there is some W_1 such that $(\forall x_i)(\forall x_j)(\langle x_i,x_j\rangle \in W_1 \Leftrightarrow x_i \in x_j)$. For $x_j \in x_i$, axiom (B1) implies that there is some W_2 such that $(\forall x_i)(\forall x_j)(\langle x_i,x_j\rangle \in W_2 \Leftrightarrow x_j \in x_i)$ and then, by Exercise 4.11(a), there is some W_3 such that $(\forall x_i)(x_j)(\langle x_i,x_j\rangle \in W_3 \Leftrightarrow x_j \in x_i)$. So, in both cases, there is some W such that $(\forall x_i)(\forall x_j)(\langle x_i,x_j\rangle \in W \Leftrightarrow \varphi(x_1,\dots,x_n,Y_1,\dots,Y_m))$. Then, by Exercise 4.11(i) with $W=X$, there is

some Z_1 such that $(\forall x_1)\dots(\forall x_{i-1})(\forall x_i)(\forall x_j)(\langle x_1,\dots,x_{i-1},x_i,x_j\rangle \in Z_1 \Leftrightarrow \varphi(x_1,\dots,x_n,Y_1,\dots,Y_m))$. Hence, by Exercise 4.11(e) with $Z_1 = X$, there exists Z_2 such that $(\forall x_1)\dots(\forall x_i)(\forall x_{i+1})\dots(\forall x_j)(\langle x_1,\dots,x_j\rangle \in Z_1 \Leftrightarrow \varphi(x_1,\dots,x_n, Y_1,\dots,Y_m))$. Then, by Exercise 4.11(d) with $Z_2 = X$, there exists Z such that $(\forall x_1)\dots(\forall x_n)(\langle x_1,\dots,x_n\rangle \in Z \Leftrightarrow \varphi(x_1,\dots,x_n,Y_1,\dots,Y_m))$. In the remaining case, $x_i \in Y_\ell$, the theorem follows by application of Exercise 4.11(f,d).

Induction step. Assume the theorem provable for all $k < r$ and assume that φ has r connectives and quantifiers.

(a) φ is $\neg\psi$. By inductive hypothesis, there is some W such that $(\forall x_1)\dots(\forall x_n)(\langle x_1,\dots,x_n\rangle \in W \Leftrightarrow \psi(x_1,\dots,x_n,Y_1,\dots,Y_m))$. Let $Z = \overline{W}$.

(b) φ *is* $\psi \Rightarrow \vartheta$. By inductive hypothesis, there are classes Z_1 and Z_2 such that $(\forall x_1)\dots(\forall x_n)(\langle x_1,\dots,x_n\rangle \in Z_1 \Leftrightarrow \psi(x_1,\dots,x_n,Y_1,\dots,Y_m))$ and $(\forall x_1)\dots(\forall x_n)(\langle x_1,\dots,x_n\rangle \in Z_2 \Leftrightarrow \vartheta(x_1,\dots,x_n,Y_1,\dots,Y_m))$. Let $Z = Z_1 \cap Z_2$.

(c) φ is $(\forall x)\psi$. By inductive hypothesis, there is some W such that $(\forall x_1)\dots(\forall x_n)(\forall x)(\langle x_1,\dots,x_n,x\rangle \in W \Leftrightarrow \psi(x_1,\dots,x_n,x,Y_1,\dots,Y_m))$. Apply Exercise 4.11(g) with $X = \overline{W}$ to obtain a class Z_1 such that $(\forall x_1)\dots(\forall x_n)(\langle x_1,\dots,x_n\rangle \in Z_1 \Leftrightarrow (\exists x)\neg\psi(x_1,\dots,x_n,x,Y_1,\dots,Y_m))$ Now let $Z = \overline{Z_1}$, noting that $(\forall x)\psi$ is equivalent to $\neg(\exists x)\neg\psi$.

Examples

1. Let $\varphi(X,Y_1,Y_2)$ be $(\exists u)(\exists v)(X = \langle u,v\rangle \wedge u \in Y_1 \wedge v \in Y_2)$. The only quantifiers in φ involve set variables. Hence, by the class existence theorem, $\vdash (\exists Z)(\forall x)(x \in Z \Leftrightarrow (\exists u)(\exists v)(x = \langle u,v\rangle \wedge u \in Y_1 \wedge v \in Y_2))$. By the extensionality principle,

$$\vdash (\exists_1 Z)(\forall x)(x \in Z \Leftrightarrow (\exists u)(\exists v)(x = \langle u,v\rangle \wedge u \in Y_1 \wedge v \in Y_2).$$

So, we can introduce a new function letter $\times$.

DEFINITION

(*Cartesian product* of Y_1 and Y_2)

$$(\forall x)(x \in Y_1 \times Y_2 \Leftrightarrow (\exists u)(\exists v)(x = \langle u,v\rangle \wedge u \in Y_1 \wedge v \in Y_2))$$

DEFINITIONS

Y^2	for $Y \times Y$	
Y^n	for $Y^{n-1} \times Y$	when $n > 2$
Rel(X)	for $X \subseteq V^2$	(X is a *relation*)†

V^2 is the class of all ordered pairs, and V^n is the class of all ordered n-tuples. In ordinary language, the word 'relation' indicates some kind of connection between objects. For example, the *parenthood relation* holds

†More precisely, Rel(X) means that X is a *binary relation*.

between parents and their children. For our purposes, we interpret the parenthood relation to be the class of all ordered pairs $\langle u, v\rangle$ such that u is a parent of v.

2. Let $\varphi(X, Y)$ be $X \subseteq Y$. By the class existence theorem and the extensionality principle, $\vdash (\exists_1 Z)(\forall x)(x \in Z \Leftrightarrow x \subseteq Y)$. Thus, there is a unique class Z that has as its members all *subsets* of Y. Z is called the *power class* of Y and is denoted $\mathscr{P}(Y)$.

DEFINITION

$(\forall x)(x \in \mathscr{P}(Y) \Leftrightarrow x \subseteq Y)$

3. Let $\varphi(X, Y)$ be $(\exists v)(X \in v \wedge v \in Y)$. By the class existence theorem and the extensionality principle, $\vdash (\exists_1 Z)(\forall x)(x \in Z \Leftrightarrow (\exists v)(x \in v \wedge v \in Y))$. Thus, there is a unique class Z that contains all members of members of Y. Z is called the *sum class* of Y and is denoted $\bigcup Y$.

DEFINITION

$(\forall x)(x \in \bigcup Y \Leftrightarrow (\exists v)(x \in v \wedge v \in Y))$

4. Let $\varphi(X)$ be $(\exists u)(X = \langle u, u\rangle)$. By the class existence theorem and the extensionality principle, there is a unique class Z such that $(\forall x)(x \in Z \Leftrightarrow (\exists u)(x = \langle u, u\rangle))$. Z is called the *identity relation* and is denoted I.

DEFINITION

$(\forall x)(x \in I \Leftrightarrow (\exists u)(x = \langle u, u\rangle))$

COROLLARY 4.5

If $\varphi(X_1, \ldots, X_n, Y_1, \ldots, Y_m)$ is a predicative wf, then

$$\vdash (\exists_1 W)(W \subseteq V^n \wedge (\forall x_1)\ldots(\forall x_n)(\langle x_1, \ldots, x_n\rangle \in W \Leftrightarrow \varphi(x_1, \ldots, x_n, Y_1, \ldots, Y_m)))$$

Proof

By Proposition 4.4, there is some Z such that $(\forall x_1)\ldots(\forall x_n)(\langle x_1, \ldots, x_n\rangle \in Z \Leftrightarrow \varphi(x_1, \ldots, x_n, Y_1, \ldots, Y_m))$. Then $W = Z \cap V^n$ satisfies the corollary, and the uniqueness follows from the extensionality principle.

DEFINITION

Given a predicative wf $\varphi(X_1, \ldots, X_n, Y_1, \ldots, Y_m)$, let $\{\langle x_1, \ldots, x_n\rangle | \varphi(x_1, \ldots, x_n, Y_1, \ldots, Y_m)\}$ denote the class of all n-tuples $\langle x_1, \ldots, x_n\rangle$ that satisfy $\varphi(x_1, \ldots, x_n, Y_1, \ldots, Y_m)$; that is,

$$(\forall u)(u \in \{\langle x_1,\ldots,x_n\rangle | \varphi(x_1,\ldots,x_n,Y_1,\ldots,Y_m)\} \Leftrightarrow$$
$$(\exists x_1)\ldots(\exists x_n)(u = \langle x_1,\ldots,x_n\rangle \wedge \varphi(x_1,\ldots,x_n,Y_1,\ldots,Y_m)))$$

This definition is justified by Corollary 4.5. In particular, when $n = 1$, $\vdash (\forall u)(u \in \{x|\varphi(x,Y_1,\ldots,Y_m)\} \Leftrightarrow \varphi(u,Y_1,\ldots,Y_m))$.

Examples

1. Take φ to be $\langle x_2,x_1\rangle \in Y$. Let $\breve{Y}$ be an abbreviation for $\{\langle x_1,x_2\rangle|\langle x_2,x_1\rangle \in Y\}$. Hence, $\breve{Y} \subseteq V^2 \wedge (\forall x_1)(\forall x_2)(\langle x_1,x_2\rangle \in \breve{Y} \Leftrightarrow \langle x_2,x_1\rangle \in Y)$. Call $\breve{Y}$ the *inverse relation* of Y.
2. Take φ to be $(\exists v)(\langle v,x\rangle \in Y)$. Let $\mathscr{R}(Y)$ stand for $\{x|(\exists v)(\langle v,x\rangle \in Y)\}$. Then $\vdash (\forall u)(u \in \mathscr{R}(Y) \Leftrightarrow (\exists v)(\langle v,x\rangle \in Y))$. $\mathscr{R}(Y)$ is called the *range* of Y. Clearly, $\vdash \mathscr{R}(Y) = \mathscr{D}(\breve{Y})$.

Notice that axioms (B1)–(B7) are special cases of the class existence theorem, Proposition 4.4. Thus, instead of the infinite number of instances of the axiom schema in Proposition 4.4, it sufficed to assume only a finite number of instances of that schema.

Exercises

4.12 Prove:

(a) $\vdash \bigcup\emptyset = \emptyset$
(b) $\vdash \bigcup\{\emptyset\} = \emptyset$
(c) $\vdash \bigcup V = V$
(d) $\vdash \mathscr{P}(V) = V$
(e) $\vdash X \subseteq Y \Rightarrow \bigcup X \subseteq \bigcup Y \wedge \mathscr{P}(X) \subseteq \mathscr{P}(Y)$
(f) $\vdash \bigcup\mathscr{P}(X) = X$
(g) $\vdash X \subseteq \mathscr{P}(\bigcup X)$
(h) $\vdash (X\cap Y) \times (W\cap Z) = (X \times W)\cap(Y \times Z)$
(i) $\vdash (X\cup Y) \times (W\cup Z) = (X \times W)\cup(X \times Z)\cup(Y \times W)\cap(Y \times Z)$
(j) $\vdash \mathscr{P}(X\cap Y) = \mathscr{P}(X)\cap\mathscr{P}(Y)$
(k) $\vdash \mathscr{P}(X)\cup\mathscr{P}(Y) \subseteq \mathscr{P}(X\cup Y)$
(l) What simple condition on X and Y is equivalent to $\mathscr{P}(X\cup Y) \subseteq \mathscr{P}(X)\cup\mathscr{P}(Y)$?
(m) $\vdash \bigcup(X\cup Y) = (\bigcup X)\cup(\bigcup Y)$
(n) $\vdash \bigcup(X\cap Y) \subseteq (\bigcup X)\cap(\bigcup Y)$
(o) $\vdash Z = \breve{Y} \Rightarrow \breve{Z} = Y\cap V^2$
(p) $\vdash \mathrm{Rel}(I) \wedge \breve{I} = I$
(q) $\vdash \mathscr{P}(\emptyset) = \{\emptyset\}$
(r) $\vdash \mathscr{P}(\{\emptyset\}) = \{\emptyset,\{\emptyset\}\}$
(s) $\vdash (\forall x)(\forall y)(x \times y \subseteq \mathscr{P}(\mathscr{P}(x\cup y)))$
(t) $\vdash \mathrm{Rel}(Y) \Rightarrow Y \subseteq \mathscr{D}(Y) \times \mathscr{R}(Y)$

Until now, although we can prove, using Proposition 4.4, the existence of a great many classes, the existence of only a few sets, such as $\emptyset$, $\{\emptyset\}$, $\{\emptyset, \{\emptyset\}\}$, and $\{\{\emptyset\}\}$, is known to us. To guarantee the existence of sets of greater complexity, we require more axioms.

AXIOM U (SUM SET)

$$(\forall x)(\exists y)(\forall u)(u \in y \Leftrightarrow (\exists v)(u \in v \wedge v \in x))$$

This axiom asserts that the sum class $\bigcup x$ of a set x is also a set, which we shall call the *sum set* of x, that is, $\vdash (\forall x)\mathrm{M}(\bigcup x)$. The sum set $\bigcup x$ is usually referred to as the *union* of all the sets in the set x and is often denoted $\bigcup_{v \in x} v$.

Exercises

4.13 Prove:

(a) $\vdash (\forall x)(\forall y)(\bigcup\{x, y\} = x \cup y)$

(b) $\vdash (\forall x)(\forall y)\mathrm{M}(x \cup y)$

(c) $\vdash (\forall x)(\bigcup\{x\} = x)$

(d) $\vdash (\forall x)(\forall y)(\bigcup\langle x, y\rangle = \{x, y\})$

4.14 Define by induction $\{x_1, \ldots, x_n\}$ to be $\{x_1, \ldots, x_{n-1}\} \cup \{x_n\}$. Prove $\vdash (\forall x_1) \ldots (\forall x_n)(\forall u)(u \in \{x_1, \ldots, x_n\} \Leftrightarrow u = x_1 \vee \ldots \vee u = x_n)$ Thus, for any sets $x_1, \ldots, x_n$, there is a set that has $x_1, \ldots, x_n$ as its only members.

Another means of generating new sets from old is the formation of the set of subsets of a given set.

AXIOM W (POWER SET)

$$(\forall x)(\exists y)(\forall u)(u \in y \Leftrightarrow u \subseteq x)$$

This axiom asserts that the power class $\mathscr{P}(x)$ of a set x is itself a set, that is, $\vdash (\forall x)\mathrm{M}(\mathscr{P}(x))$.

A much more general way to produce sets is the following *axiom of subsets.*

AXIOM S (SUBSETS)

$$(\forall x)(\forall Y)(\exists z)(\forall u)(u \in z \Leftrightarrow u \in x \wedge u \in Y)$$

COROLLARY 4.6

(a) $\vdash (\forall x)(\forall Y)\mathrm{M}(x \cap Y)$ (The intersection of a set and a class is a set.)
(b) $\vdash (\forall x)(\forall Y)(Y \subseteq x \Rightarrow \mathrm{M}(Y))$ (A subclass of a set is a set.)
(c) For any predicative wf $\mathscr{B}(y)$, $\vdash (\forall x)\mathrm{M}(\{y|y \in x \wedge \mathscr{B}(y)\})$.

Proof

(a) By axiom S, there is a set z such that $(\forall u)(u \in z \Leftrightarrow u \in x \wedge u \in Y)$, which implies $(\forall u)(u \in z \Leftrightarrow u \in x \cap Y)$. Thus, $z = x \cap Y$ and, therefore, $x \cap Y$ is a set.
(b) If $Y \subseteq x$, then $x \cap Y = Y$ and the result follows by part (a).
(c) Let $Y = \{y|y \in x \wedge \mathscr{B}(y)\}$[†]. Since $Y \subseteq x$, part (b) implies that Y is a set.

Exercise

4.15 Prove:
(a) $\vdash (\forall x)(\mathrm{M}(\mathscr{D}(x)) \wedge \mathrm{M}(\mathscr{R}(x)))$.
(b) $\vdash (\forall x)(\forall y)\mathrm{M}(x \times y)$. [*Hint:* Exercise 4.12(s).]
(c) $\vdash \mathrm{M}(\mathscr{D}(Y)) \wedge \mathrm{M}(\mathscr{R}(Y)) \wedge \mathrm{Rel}(Y) \Rightarrow \mathrm{M}(Y)$. [*Hint:* Exercise 4.12(t).]
(d) $\vdash \mathrm{Pr}(Y) \wedge Y \subseteq X \Rightarrow \mathrm{Pr}(X)$.

On the basis of axiom S, we can show that the intersection of any non-empty class of sets is a set.

DEFINITION

$\bigcap X$ for $\{y|(\forall x)(x \in X \Rightarrow y \in x)\}$ (intersection)

PROPOSITION 4.7

(a) $\vdash (\forall x)(x \in X \Rightarrow \bigcap X \subseteq x)$
(b) $\vdash X \neq \emptyset \Rightarrow \mathrm{M}(\bigcap X)$
(c) $\vdash \bigcap \emptyset = V$

Proof

(a) Assume $u \in X$. Consider any y in $\bigcap X$. Then $(\forall x)(x \in X \Rightarrow y \in x)$. Hence, $y \in u$. Thus, $\bigcap X \subseteq u$.
(b) Assume $X \neq \emptyset$. Let $x \in X$. By part (a), $\bigcap X \subseteq x$. Hence, by Corollary 4.6(b), $\bigcap X$ is a set.

[†]More precisely, the wf $Y \in X \wedge \mathscr{B}(Y)$ is predicative, so that the class existence theorem yields a class $\{y|y \in X \wedge \mathscr{B}(y)\}$. In our case, X is a set x.

(c) Since $\vdash (\forall x)(x \notin \emptyset)$, $\vdash (\forall y)(\forall x)(x \in \emptyset \Rightarrow y \in x)$, from which we obtain $\vdash (\forall y)(y \in \bigcap \emptyset)$. From $\vdash (\forall y)(y \in V)$ and the extensionality principle, $\vdash \bigcap \emptyset = V$.

Exercise

4.16 Prove:

(a) $\vdash \bigcap\{x, y\} = x \cap y$

(b) $\vdash \bigcap\{x\} = x$

(c) $\vdash X \subseteq Y \Rightarrow \bigcap Y \subseteq \bigcap X$

A stronger axiom than axiom S will be necessary for the full development of set theory. First, a few definitions are convenient.

DEFINITIONS

$\mathrm{Fnc}(X)$ for $\mathrm{Rel}(X) \wedge (\forall x)(\forall y)(\forall z)(\langle x, y\rangle \in X \wedge \langle x, z\rangle \in X \Rightarrow y = z)$ (X is a *function*)

$X : Y \to Z$ for $\mathrm{Fnc}(X) \wedge \mathscr{D}(X) = Y \wedge \mathscr{R}(X) \subseteq Z$ (X is a *function from Y into Z*)

$Y \restriction X$ for $X \cap (Y \times V)$ (*restriction* of X to the domain Y)

$\mathrm{Fnc}_1(X)$ for $\mathrm{Fnc}(X) \wedge \mathrm{Fnc}(\breve{X})$ (X is a *one – one* function)

$$X`Y = \begin{cases} z & \text{if } (\forall u)(\langle Y, u\rangle \in X \Leftrightarrow u = z) \\ \emptyset & \text{otherwise} \end{cases}$$

$X``Y = \mathscr{R}(Y \restriction X)$

If there is a unique z such that $\langle y, z\rangle \in X$, then $z = X`y$; otherwise, $X`y = \emptyset$. If X is a function and y is a set in its domain, $X`y$ is the value of the function applied to y. If X is a function, $X``Y$ is the range of X restricted to Y.†

Exercise

4.17 Prove:

(a) $\vdash \mathrm{Fnc}(X) \wedge y \in \mathscr{D}(X) \Rightarrow (\forall z)(X`y = z \Leftrightarrow \langle y, z\rangle \in X)$

(b) $\vdash \mathrm{Fnc}(X) \wedge Y \subseteq \mathscr{D}(X) \Rightarrow \mathrm{Fnc}(Y \restriction X) \wedge \mathscr{D}(Y \restriction X) = Y \wedge (\forall y)(y \in Y \Rightarrow X`y = (Y \restriction X)`y)$

(c) $\vdash \mathrm{Fnc}(X) \Rightarrow [\mathrm{Fnc}_1(X) \Leftrightarrow (\forall y)(\forall z)(y \in \mathscr{D}(X) \wedge z \in \mathscr{D}(X) \wedge y \neq z \Rightarrow X`y \neq X`z)]$

(d) $\vdash \mathrm{Fnc}(X) \wedge Y \subseteq \mathscr{D}(X) \Rightarrow (\forall z)(z \in X``Y \Leftrightarrow (\exists y)(y \in Y \wedge X`y = z))$

†In traditional set-theoretic notation, if F is a function and y is in its domain, $F`y$ is written as $F(y)$, and if Y is included in the domain of F, $F``Y$ is sometimes written as $F[Y]$.

AXIOM R (REPLACEMENT)

$$\mathrm{Fnc}(Y) \Rightarrow (\forall x)(\exists y)(\forall u)(u \in y \Leftrightarrow (\exists v)(\langle v, u\rangle \in Y \wedge v \in x))$$

Axiom R asserts that, if Y is a function and x is a set, then the class of second components of ordered pairs in Y whose first components are in x is a set (or, equivalently, $\mathscr{R}(x\{Y)$ is a set).

Exercises

4.18 Show that, in the presence of the other axioms, the replacement axiom (R) implies the axiom of subsets (S).
4.19 Prove $\vdash \mathrm{Fnc}(Y) \Rightarrow (\forall x)\mathrm{M}(Y\text{“}x))$.
4.20 Show that axiom R is equivalent to the wf
$\mathrm{Fnc}(Y) \wedge \mathrm{M}(\mathscr{D}(Y)) \Rightarrow \mathrm{M}(\mathscr{R}(Y))$.
4.21 Show that, in the presence of all axioms except R and S, axiom R is equivalent to the conjunction of axiom S and the wf
$\mathrm{Fnc}_1(Y) \wedge \mathrm{M}(\mathscr{D}(Y)) \Rightarrow \mathrm{M}(\mathscr{R}(Y))$.

To ensure the existence of an infinite set, we add the following axiom.

AXIOM I (AXIOM OF INFINITY)

$$(\exists x)(\emptyset \in x \wedge (\forall u)(u \in x \Rightarrow u \cup \{u\} \in x))$$

Axiom I states that there is a set x that contains $\emptyset$ and such that, whenever a set u belongs to x, then $u \cup \{u\}$ also belongs to x. Hence, for such a set x, $\{\emptyset\} \in x$, $\{\emptyset, \{\emptyset\}\} \in x$, $\{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\} \in x$, and so on. If we let 1 stand for $\{\emptyset\}$, 2 for $\{\emptyset, 1\}$, 3 for $\{\emptyset, 1, 2\}, \ldots, n$ for $\{\emptyset, 1, 2, \ldots, n-1\}$, etc., then, for all ordinary integers $n \geq 0$, $n \in x$, and $\emptyset \neq 1$, $\emptyset \neq 2$, $1 \neq 2$, $\emptyset \neq 3$, $1 \neq 3$, $2 \neq 3, \ldots$.

Exercise

4.22 (a) Prove that any wf that implies $(\exists X)\mathrm{M}(X)$ would, together with axiom S, imply axiom N.
(b) Show that axiom I is equivalent to the following sentence (I^*):

$$(\exists x)((\exists y)(y \in x \wedge (\forall u)(u \notin y)) \wedge (\forall u)(u \in x \Rightarrow u \cup \{u\} \in x))$$

Then prove that (I^*) implies axiom N. (Hence, if we assumed (I^*) instead of (I), axiom N would become superfluous.)

This completes the list of axioms of NBG, and we see that NBG has only a finite number of axioms – namely, axiom T, axiom P (pairing), axiom N (null set), axiom U (sum set), axiom W (power set), axiom S (subsets), axiom R (replacement), axiom I (infinity), and the seven class existence axioms (B1)–(B7). We have also seen that axiom S is provable from the other axioms; it has been included here because it is of interest in the study of certain weaker subtheories of NBG.

Let us verify now that the usual argument for Russell's paradox does not hold in NBG. By the class existence theorem, there is a class $Y = \{x | x \notin x\}$. Then $(\forall x)(x \in Y \Leftrightarrow x \notin x)$. In unabbreviated notation this becomes $(\forall X)(\mathrm{M}(X) \Rightarrow (X \in Y \Leftrightarrow X \notin X))$. Assume $\mathrm{M}(Y)$. Then $Y \in Y \Leftrightarrow Y \notin Y$, which, by the tautology $(A \Leftrightarrow \neg A) \Rightarrow (A \wedge \neg A)$, yields $Y \in Y \wedge Y \notin Y$. Hence, by the derived rule of proof by contradiction, we obtain $\vdash \neg\mathrm{M}(Y)$. Thus, in NBG, the argument for Russell's paradox merely shows that Russell's class Y is a proper class, not a set. NBG will avoid the paradoxes of Cantor and Burali-Forti in a similar way.

Exercise

4.23 Prove $\vdash \neg\mathrm{M}(V)$, that is, the universal class V is not a set. [*Hint:* Apply Corollary 4.6(b) with Russell's class Y.]

4.2 ORDINAL NUMBERS

Let us first define some familiar notions concerning relations.

DEFINITIONS

X Irr Y for $\mathrm{Rel}(X) \wedge (\forall y)(y \in Y \Rightarrow \langle y, y\rangle \notin X)$
(X is an *irreflexive* relation on Y)

X Tr Y for $\mathrm{Rel}(X) \wedge (\forall u)(\forall v)(\forall w)([u \in Y \wedge v \in Y \wedge w \in Y \wedge \langle u, v\rangle \in X \wedge \langle v, w\rangle \in X] \Rightarrow \langle u, w\rangle \in X)$
(X is a *transitive* relation on Y)

X Part Y for $(X \text{ Irr } Y) \wedge (X \text{ Tr } Y)$ (X *partially orders* Y)

X Con Y for $\mathrm{Rel}(X) \wedge (\forall u)(\forall v)([u \in Y \wedge v \in Y \wedge u \neq v] \Rightarrow \langle u, v\rangle \in X \vee \langle v, u\rangle \in X)$
(X is a *connected* relation on Y)

X Tot Y for $(X \text{ Irr } Y) \wedge (X \text{ Tr } Y \wedge (X \text{ Con } Y)$ (X *totally orders* Y)

X We Y for $(X \text{ Irr } Y) \wedge (\forall Z)([Z \subseteq Y \wedge Z \neq \emptyset] \Rightarrow (\exists y)(y \in Z \wedge (\forall v)(v \in Z \wedge v \neq y \Rightarrow \langle y, v\rangle \in X \wedge \langle v, y\rangle \notin X)))$

(*X well-orders Y*, that is, the relation X is irreflexive on Y and every non-empty subclass of Y has a least element with respect to X)

Exercises

4.24 Prove $\vdash X$ We $Y \Rightarrow X$ Tot Y. [*Hint :* To show X Con Y, let $x \in Y \wedge y \in Y \wedge x \neq y$. Then $\{x,y\}$ has a least element, say x. Then $\langle x,y\rangle \in X$. To show X Tr Y, assume $x \in Y \wedge y \in Y \wedge z \in Y \wedge \langle x,y\rangle \in X \wedge \langle y,z\rangle \in X$. Then $\{x,y,z\}$ has a least element, which must be x.]

4.25 Prove $\vdash X$ We $Y \wedge Z \subseteq Y \Rightarrow X$ We Z.

Examples (from intuitive set theory)

1. The relation $<$ on the set P of positive integers well-orders P.
2. The relation $<$ on the set of all integers totally orders, but does not well-order, this set. The set has no least element.
3. The relation $\subset$ on the set W of all subsets of the set of integers partially orders W but does not totally order W. For example, $\{1\} \not\subset \{2\}$ and $\{2\} \not\subset \{1\}$.

DEFINITION

$\text{Simp}(Z, W_1, W_2)$ for

$(\exists x_1)(\exists x_2)(\exists r_1)(\exists r_2)(\text{Rel}(r_1) \wedge \text{Rel}(r_2) \wedge W_1 = \langle r_1, x_1\rangle \wedge W_2 = \langle r_2, x_2\rangle$

$\wedge\ \text{Fnc}_1(Z) \wedge \mathscr{D}(Z) = x_1 \wedge \mathscr{R}(Z) = x_2 \wedge (\forall u)(\forall v)(u \in x_1 \wedge v \in x_1 \Rightarrow$

$(\langle u,v\rangle \in r_1 \Leftrightarrow \langle Z`u, Z`v\rangle \in r_2)))$

(Z is a *similarity mapping* of the relation r_1 on x_1 onto the relation r_2 on x_2.)

DEFINITION

$\text{Sim}(W_1, W_2)$ for $(\exists z)\text{Simp}(z, W_1, W_2)$

(W_1 and W_2 are *similar ordered structures*)

Example

Let r_1 be the less-than relation $<$ on the set A of non-negative integers $\{0,1,2,\ldots\}$, and let r_2 be the less-than relation $<$ on the set B of positive integers $\{1,2,3,\ldots\}$. Let z be the set of all ordered pairs $\langle x, x+1\rangle$ for $x \in A$. Then z is a similarity mapping of $\langle r_1, A\rangle$ onto $\langle r_2, B\rangle$.

DEFINITION

$X_1 \circ X_2$ for $\{\langle u,v\rangle | (\exists z)(\langle u,z\rangle \in X_2 \wedge \langle z,v\rangle \in X_1)\}$

(the *composition* of X_2 and X_1)

Exercises

4.26 Prove:

(a) $\vdash \mathrm{Simp}(Z,X,Y) \Rightarrow \mathrm{M}(Z) \wedge \mathrm{M}(X) \wedge \mathrm{M}(Y)$
(b) $\vdash \mathrm{Simp}(Z,X,Y) \Rightarrow \mathrm{Simp}(\breve{Z},Y,X)$

4.27

(a) Prove: $\vdash \mathrm{Rel}(X_1) \wedge \mathrm{Rel}(X_2) \Rightarrow \mathrm{Rel}(X_1 \circ X_2)$
(b) Let X_1 and X_2 be the parent and brother relations on the set of human beings. What are the relations $X_1 \circ X_1$ and $X_1 \circ X_2$?
(c) Prove: $\vdash \mathrm{Fnc}(X_1) \wedge \mathrm{Fnc}(X_2) \Rightarrow \mathrm{Fnc}(X_1 \circ X_2)$
(d) Prove: $\vdash \mathrm{Fnc}_1(X_1) \wedge \mathrm{Fnc}_1(X_2) \Rightarrow \mathrm{Fnc}_1(X_1 \circ X_2)$
(e) Prove: $\vdash (X_1 : Z \to W \wedge X_2 : Y \to Z) \Rightarrow X_1 \circ X_2 : Y \to W$

DEFINITIONS

$\mathrm{Fld}(X)$ for $\mathscr{D}(X) \cup \mathscr{R}(X)$	(the *field* of X)
$\mathrm{TOR}(X)$ for $\mathrm{Rel}(X) \wedge (X \text{ Tot } (\mathrm{Fld}(X)))$	(X is a *total order*)
$\mathrm{WOR}(X)$ for $\mathrm{Rel}(X) \wedge (X \text{ We } (\mathrm{Fld}(X)))$	(X is a *well – ordering relation*)

Exercise

4.28 Prove:

(a) $\vdash \mathrm{Sim}(W_1, W_2) \Rightarrow \mathrm{Sim}(W_2, W_1)$
(b) $\vdash \mathrm{Sim}(W_1, W_2) \wedge \mathrm{Sim}(W_2, W_3) \Rightarrow \mathrm{Sim}(W_1, W_3)$
(c) $\vdash \mathrm{Sim}(\langle X, \mathrm{Fld}(X)\rangle, \langle Y, \mathrm{Fld}(Y)\rangle) \Rightarrow (\mathrm{TOR}(X) \Leftrightarrow \mathrm{TOR}(Y)) \wedge (\mathrm{WOR}(X) \Leftrightarrow \mathrm{WOR}(Y))$

If x is a total order, then the class of all total orders similar to x is called the *order type* of x. We are especially interested in the order types of well-ordering relations, but, since it turns out that all order types are proper classes (except the order type $\{\emptyset\}$ of $\emptyset$), it will be convenient to find a class W of well-ordered structures such that every well-ordering is similar to a unique member of W. This leads us to the study of ordinal numbers.

DEFINITIONS

E for $\{\langle x,y\rangle \mid x \in y\}$	(the *membership relation*)
$\mathrm{Trans}(X)$ for $(\forall u)(u \in X \Rightarrow u \subseteq X)$	(X is *transitive*)
$\mathrm{Sect}_Y(X,Z)$ for	

$$Z \subseteq X \wedge (\forall u)(\forall v)([u \in X \wedge v \in Z \wedge \langle u,v\rangle \in Y] \Rightarrow u \in Z)$$

(Z is a *Y-section* of X, that is, Z is a subclass of X and every

member of X that Y-precedes a member of Z is also a member of Z.)

$\mathrm{Seg}_Y(X,W)$ for $\{x|x \in X \wedge \langle x, W\rangle \in Y\}$(the *Y-segment* of X determined by W, that is, the class of all members of X that Y-precede W)

Exercises

4.29 Prove:

(a) $\vdash \mathrm{Trans}(X) \Leftrightarrow (\forall u)(\forall v)(v \in u \wedge u \in X \Rightarrow v \in X)$

(b) $\vdash \mathrm{Trans}(X) \Leftrightarrow \bigcup X \subseteq X$

(c) $\vdash \mathrm{Trans}(\emptyset)$

(d) $\vdash \mathrm{Trans}(\{\emptyset\})$

(e) $\vdash \mathrm{Trans}(X) \wedge \mathrm{Trans}(Y) \Rightarrow \mathrm{Trans}(X \cup Y) \wedge \mathrm{Trans}(X \cap Y)$

(f) $\vdash \mathrm{Trans}(X) \Rightarrow \mathrm{Trans}(\bigcup X)$

(g) $\vdash (\forall u)(u \in X) \Rightarrow \mathrm{Trans}(u)) \Rightarrow \mathrm{Trans}(\bigcup X)$

4.30 Prove:

(a) $\vdash (\forall u)[\mathrm{Seg}_E(X,u) = X \cap u \wedge \mathrm{M}(\mathrm{Seg}_E(X,u))]$

(b) $\vdash \mathrm{Trans}(X) \Leftrightarrow (\forall u)(u \in X \Rightarrow \mathrm{Seg}_E(X,u) = u)$

(c) $\vdash$ E We $X \wedge \mathrm{Sect}_E(X,Z) \wedge Z \neq X \Rightarrow (\exists u)(u \in X \wedge Z = \mathrm{Seg}_E(X,u))$

DEFINITIONS

$\mathrm{Ord}(X)$ for E We $X \wedge \mathrm{Trans}(X)$ (X is an *ordinal class* if and only if the $\in$-relation well-orders X and any member of X is a subset of X)

On for $\{x|\mathrm{Ord}(x)\}$ (The class of *ordinal numbers*)

Thus, $\vdash (\forall x)(x \in On \Leftrightarrow \mathrm{Ord}(x))$. An ordinal class that is a set is called an ordinal number, and *On* is the class of all ordinal numbers. Notice that a wf $x \in On$ is equivalent to a predicative wf – namely, the conjunction of the following wfs:

(a) $(\forall u)(u \in x \Rightarrow u \notin u)$

(b) $(\forall u)(u \subseteq x \wedge u \neq \emptyset \Rightarrow (\exists v)(v \in u \wedge (\forall w)(w \in u \wedge w \neq v \Rightarrow v \in w \wedge w \notin v)))$

(c) $(\forall u)(u \in x \Rightarrow u \subseteq x)$

(The conjunction of (a) and (b) is equivalent to E We x, and (c) is $\mathrm{Trans}(x)$.) In addition, any wf $On \in Y$ can be replaced by the wf $(\exists y)(y \in Y \wedge (\forall z)(z \in y \Leftrightarrow z \in On))$. Hence, any wf that is predicative except for the presence of '*On*' is equivalent to a predicative wf and therefore can be used in connection with the class existence theorem.

Exercise

4.31 Prove: (a) $\vdash \emptyset \in On$. (b) $\vdash 1 \in On$, where 1 stands for $\{\emptyset\}$.

We shall use lower-case Greek letters $\alpha, \beta, \gamma, \delta, \tau, \ldots$ as restricted variables for ordinal numbers. Thus, $(\forall\alpha)\mathscr{B}(\alpha)$ stands for $(\forall x)(x \in On \Rightarrow \mathscr{B}(x))$, and $(\exists\alpha)\mathscr{B}(\alpha)$ stands for $(\exists x)(x \in On \wedge \mathscr{B}(x))$.

PROPOSITION 4.8

(a) $\vdash \mathrm{Ord}(X) \Rightarrow (X \notin X \wedge (\forall u)(u \in X \Rightarrow u \notin u))$
(b) $\vdash \mathrm{Ord}(X) \wedge Y \subset X \wedge \mathrm{Trans}(Y) \Rightarrow Y \in X$
(c) $\vdash \mathrm{Ord}(X) \wedge \mathrm{Ord}(Y) \Rightarrow (Y \subset X \Leftrightarrow Y \in X)$
(d) $\vdash \mathrm{Ord}(X) \wedge \mathrm{Ord}(Y) \Rightarrow [(X \in Y \vee X = Y \vee Y \in X) \wedge \neg(X \in Y \wedge Y \in X) \wedge \neg(X \in Y \vee X = Y)]$
(e) $\vdash \mathrm{Ord}(X) \wedge Y \in X \Rightarrow Y \in On$
(f) $\vdash$ E We On
(g) $\vdash \mathrm{Ord}(On)$
(h) $\vdash \neg\mathrm{M}(On)$
(i) $\vdash \mathrm{Ord}(X) \Rightarrow X = On \vee X \in On$
(j) $\vdash y \subseteq On \wedge \mathrm{Trans}(y) \Rightarrow y \in On$
(k) $\vdash x \in On \wedge y \in On \Rightarrow (x \subseteq y \vee y \subseteq x)$

Proof

(a) If $\mathrm{Ord}(X)$, then E is irreflexive on X; so, $(\forall u)(u \in X \Rightarrow u \notin u)$; and, if $X \in X$, $X \notin X$. Hence, $X \notin X$.
(b) Assume $\mathrm{Ord}(X) \wedge Y \subset X \wedge \mathrm{Trans}(Y)$. It is easy to see that Y is a proper E-section of X. Hence, by Exercise 4.30(b,c), $Y \in X$.
(c) Assume $\mathrm{Ord}(X) \wedge \mathrm{Ord}(Y)$. If $Y \in X$, then $Y \subseteq X$, since X is transitive; but $Y \neq X$ by (a); so, $Y \subset X$. Conversely, if $Y \subset X$, then, since Y is transitive, we have $Y \in X$ by (b).
(d) Assume $\mathrm{Ord}(X) \wedge \mathrm{Ord}(Y) \wedge X \neq Y$. Now, $X \cap Y \subseteq X$ and $X \cap Y \subseteq Y$. Since X and Y are transitive, so is $X \cap Y$. If $X \cap Y \subset X$ and $X \cap Y \subset Y$, then, by (b), $X \cap Y \in X$ and $X \cap Y \in Y$; hence, $X \cap Y \in X \cap Y$, contradicting the irreflexivity of E on X. Hence, either $X \cap Y = X$ or $X \cap Y = Y$; that is, $X \subseteq Y$ or $Y \subseteq X$. But $X \neq Y$. Hence, by (c), $X \in Y$ or $Y \in X$. Also, if $X \in Y$ and $Y \in X$, then, by (c), $X \subset Y$ and $Y \subset X$, which is impossible. Clearly, $X \in Y \wedge X = Y$ is impossible, by (a).
(e) Assume $\mathrm{Ord}(X) \wedge Y \in X$. We must show E We Y and $\mathrm{Trans}(Y)$. Since $Y \in X$ and $\mathrm{Trans}(X)$, $Y \subset X$. Hence, since E We X, E We Y. Moreover, if $u \in Y$ and $v \in u$, then, by $\mathrm{Trans}(X)$, $v \in X$. Since E Con X and $Y \in X \wedge v \in X$, then $v \in Y \vee v = Y \vee Y \in v$. If either $v = Y$ or $Y \in v$,

then, since E Tr X and $u \in Y \wedge v \in u$, we would have $u \in u$, contradicting (a). Hence $v \in Y$. So, if $u \in Y$, then $u \subseteq Y$, that is, Trans(Y).

(f) By (a), E Irr On. Now assume $X \subseteq On \wedge X \neq \emptyset$. Let $\alpha \in X$. If α is the least element of X, we are done. (By *least element* of X we mean an element v in X such that $(\forall u)(u \in X \wedge u \neq v \Rightarrow v \in u)$.) If not, then E We α and $X \cap \alpha \neq \emptyset$; let β be the least element of $X \cap \alpha$. It is obvious, using (d), that β is the least element of X.

(g) We must show E We On and Trans(On). The first part is (f). For the second, if $u \in On$ and $v \in u$, then, by (e), $v \in On$. Hence, Trans(On).

(h) If M(On), then, by (g), $On \in On$, contradicting (a).

(i) Assume Ord(X). Then $X \subseteq On$. If $X \neq On$, then, by (c), $X \in On$.

(j) Substitute On for X and y for Y in (b). By (h), $y \subset On$.

(k) Use parts (d) and (c).

We see from Proposition 4.8(i) that the only ordinal class that is not an ordinal number is the class On itself.

DEFINITIONS

$x <_o y$ for $x \in On \wedge y \in On \wedge x \in y$

$x \leqslant_o y$ for $y \in On \wedge (x = y \vee x <_o y)$

Thus, for ordinals, $<_o$ is the same as $\in$; so, $<_o$ well-orders On. In particular, from Proposition 4.8(e) we see that any ordinal x is equal to the set of smaller ordinals.

PROPOSITION 4.9 (TRANSFINITE INDUCTION)

$$\vdash (\forall \beta)[(\forall \alpha)(\alpha \in \beta \Rightarrow \alpha \in X) \Rightarrow \beta \in X] \Rightarrow On \subseteq X$$

(If, for every β, whenever all ordinals less than β are in X, β must also be in X, then all ordinals are in X.)

Proof

Assume $(\forall \beta)[(\forall \alpha)(\alpha \in \beta \Rightarrow \alpha \in X) \Rightarrow \beta \in X]$. Assume there is an ordinal in $On - X$. Then, since On is well-ordered by E, there is a least ordinal β in $On - X$. Hence, all ordinals less than β are in X. So, by hypothesis, β is in X, which is a contradiction.

Proposition 4.9 is used to prove that all ordinals have a given property $\mathscr{B}(\alpha)$. We let $X = \{x | \mathscr{B}(x) \wedge x \in On\}$ and show that $(\forall \beta)[(\forall \alpha)(\alpha \in \beta \Rightarrow \mathscr{B}(\alpha)) \Rightarrow \mathscr{B}(\beta)]$.

DEFINITION

x' for $x \cup \{x\}$

PROPOSITION 4.10

(a) $\vdash (\forall x)(x \in On \Leftrightarrow x' \in On)$
(b) $\vdash (\forall \alpha)\neg(\exists \beta)(\alpha <_o \beta <_o \alpha')$
(c) $\vdash (\forall \alpha)(\forall \beta)(\alpha' = \beta' \Rightarrow \alpha = \beta)$

Proof

(a) $x \in x'$. Hence, if $x' \in On$, then $x \in On$ by Proposition 4.8(e). Conversely, assume $x \in On$. We must prove E We $(x \cup \{x\})$ and Trans$(x \cup \{x\})$. Since E We x and $x \notin x$, E Irr $(x \cup \{x\})$. Also, if $y \neq \emptyset \wedge y \subseteq x \cup \{x\}$, then either $y = \{x\}$, in which case the least element of y is x, or $y \cap x \neq \emptyset$ and the least element of $y \cap x$ is then the least element of y. Hence, E We $(x \cup \{x\})$. In addition, if $y \in x \cup \{x\}$ and $u \in y$, then $u \in x$. Thus, Trans$(x \cup \{x\})$.
(b) Assume $\alpha <_o \beta <_o \alpha'$. Then, $\alpha \in \beta \wedge \beta \in \alpha'$. Since $\alpha \in \beta, \beta \notin \alpha$ and $\beta \neq \alpha$ by Proposition 4.8(d), contradicting $\beta \in \alpha'$.
(c) Assume $\alpha' = \beta'$. Then $\beta <_o \alpha'$ and, by part (b), $\beta \leqslant_o \alpha$.
Similarly, $\alpha \leqslant_o \beta$. Hence, $\alpha = \beta$.

Exercise

4.32 Prove: $\vdash (\forall \alpha)(\alpha \subset \alpha')$

DEFINITIONS

Suc(X) for $X \in On \wedge (\exists \alpha)(X = \alpha')$	(X is a *successor ordinal*)
K_1 for $\{x \mid x = \emptyset \vee \text{Suc}(x)\}$	(the class of *ordinals of the first kind*)
ω for $\{x \mid x \in K_1 \wedge \forall u)(u \in x \Rightarrow u \in K_1)\}$	(ω is the class of all ordinals α of the first kind such that all ordinals smaller than α are also of the first kind)

Example
$\vdash \emptyset \in \omega \wedge 1 \in \omega$. (Recall that $1 = \{\emptyset\}$.)

PROPOSITION 4.11

(a) $\vdash (\forall \alpha)(\alpha \in \omega \Leftrightarrow \alpha' \in \omega)$
(b) $\vdash M(\omega)$

(c) $\vdash \emptyset \in X \wedge (\forall u)(u \in X \Rightarrow u' \in X) \Rightarrow \omega \subseteq X$
(d) $\vdash (\forall \alpha)(\alpha \in \omega \wedge \beta <_o \alpha \Rightarrow \beta \in \omega)$

Proof

(a) Assume $\alpha \in \omega$. Since $\mathrm{Suc}(\alpha')$, $\alpha' \in \mathrm{K}_1$. Also, if $\beta \in \alpha'$, then $\beta \in \alpha$ or $\beta = \alpha$. Hence, $\beta \in \mathrm{K}_1$. Thus, $\alpha' \in \omega$. Conversely, if $\alpha' \in \omega$, then, since $\alpha \in \alpha'$ and $(\forall \beta)(\beta \in \alpha \Rightarrow \beta \in \alpha')$, it follows that $\alpha \in \omega$.
(b) By the axiom of infinity (I), there is a set x such that $\emptyset \in x$ and $(\forall u)(u \in x \Rightarrow u' \in x)$. We shall prove $\omega \subseteq x$. Assume not. Let α be the least ordinal in $\omega - x$. Clearly, $\alpha \neq \emptyset$, since $\emptyset \in x$. Hence, $\mathrm{Suc}(\alpha)$. So, $(\exists \beta)(\alpha = \beta')$. Let δ be an ordinal such that $\alpha = \delta'$. Then $\delta <_o \alpha$ and, by part (a), $\delta \in \omega$. Therefore, $\delta \in x$. Hence, $\delta' \in x$. But $\alpha = \delta'$. Therefore, $\alpha \in x$, which yields a contradiction. Thus, $\omega \subseteq x$. So, $\mathrm{M}(\omega)$ by Corollary 4.6(b).
(c) This is proved by a procedure similar to that used for part (b).
(d) This is left as an exercise.

The elements of ω are called *finite ordinals*. We shall use the standard notation: 1 for $\emptyset'$, 2 for $1'$, 3 for $2'$, and so on. Thus, $\emptyset \in \omega, 1 \in \omega, 2 \in \omega, 3 \in \omega, \ldots$.

The non-zero ordinals that are not successor ordinals are called *limit ordinals*.

DEFINITION

$\mathrm{Lim}(x)$ for $x \in On \wedge x \notin \mathrm{K}_1$

Exercise

4.33 Prove:
(a) $\vdash \mathrm{Lim}(\omega)$
(b) $\vdash (\forall \alpha)(\forall \beta)(\mathrm{Lim}(\alpha) \wedge \beta <_o \alpha \Rightarrow \beta' <_o \alpha)$.

PROPOSITION 4.12

(a) $\vdash (\forall x)(x \subseteq On \Rightarrow [\bigcup x \in On \wedge (\forall \alpha)(\alpha \in x \Rightarrow \alpha \leqslant_o \bigcup x) \wedge (\forall \beta)((\forall \alpha)(\alpha \in x \Rightarrow \alpha \leqslant_o \beta) \Rightarrow \bigcup x \leqslant_o \beta)])$. (If x is a set of ordinals, then $\bigcup x$ is an ordinal that is the least upper bound of x.)
(b) $\vdash (\forall x)(x \subseteq On \wedge x \neq \emptyset \wedge (\forall \alpha)(\alpha \in x \Rightarrow (\exists \beta)(\beta \in x \wedge \alpha <_o \beta))]$ $\Rightarrow \mathrm{Lim}(\bigcup x))$. (If x is a non-empty set of ordinals without a maximum, then $\bigcup x$ is a limit ordinal.)

Proof

(a) Assume $x \subseteq On$. $\bigcup x$, as a set of ordinals, is well-ordered by E. Also, if $\alpha \in \bigcup x \wedge \beta \in \alpha$, then there is some γ with $\gamma \in x$ and $\alpha \in \gamma$. Then $\beta \in \alpha \wedge \alpha \in \gamma$; since every ordinal is transitive, $\beta \in \gamma$. So, $\beta \in \bigcup x$. Hence, $\bigcup x$ is transitive and, therefore, $\bigcup x \in On$. In addition, if $\alpha \in x$, then $\alpha \subseteq \bigcup x$; so, $\alpha \leqslant_o \bigcup x$, by Proposition 4.8(c). Assume now that $(\forall\alpha)(\alpha \in x \Rightarrow \alpha \leqslant_o \beta)$. Clearly, if $\delta \in \bigcup x$, then there is some γ such that $\delta \in \gamma \wedge \gamma \in x$. Hence, $\gamma \leqslant_o \beta$ and so, $\delta <_o \beta$. Therefore, $\bigcup x \subseteq \beta$ and, by Proposition 4.8(c), $\bigcup x \leqslant_o \beta$.

(b) Assume $x \subseteq On \wedge x \neq \emptyset \wedge (\forall\alpha)(\alpha \in x \Rightarrow (\exists\beta)(\beta \in x \wedge \alpha <_o \beta))$. If $\bigcup x = \emptyset$, then $\alpha \in x$ implies $\alpha = \emptyset$. So, $x = \emptyset$ or $x = 1$, which contradicts our assumption. Hence, $\bigcup x \neq \emptyset$. Assume $\mathrm{Suc}(\bigcup x)$. Then $\bigcup x = \gamma'$ for some γ. By part (a), $\bigcup x$ is a least upper bound of x. Therefore, γ is not an upper bound of x; there is some δ in x with $\gamma <_o \delta$. But then $\delta = \bigcup x$, since $\bigcup x$ is an upper bound of x. Thus, $\bigcup x$ is a maximum element of x, contradicting our hypothesis. Hence, $\neg\mathrm{Suc}(\bigcup x)$, and $\mathrm{Lim}(\bigcup x)$ is the only possibility left.

Exercise

4.34 Prove:

(a) $\vdash (\forall\alpha)([\mathrm{Suc}(\alpha) \Rightarrow (\bigcup\alpha)' = \alpha] \wedge [\mathrm{Lim}(\alpha) \Rightarrow \bigcup\alpha = \alpha])$.

(b) If $\emptyset \neq x \subseteq On$, then $\bigcap x$ is the least ordinal in x.

We can now state and prove another form of transfinite induction.

PROPOSITION 4.13 (TRANSFINITE INDUCTION: SECOND FORM)

(a) $\vdash [\emptyset \in X \wedge (\forall\alpha)(\alpha \in X \Rightarrow \alpha' \in X) \wedge (\forall\alpha)(\mathrm{Lim}(\alpha) \wedge (\forall\beta)(\beta <_o \alpha \Rightarrow \beta \in X) \Rightarrow \alpha \in X)] \Rightarrow On \subseteq X$.

(b) (*Induction up to* δ.) $\vdash [\emptyset \in X \wedge (\forall\alpha)(\alpha <_o \delta \wedge \alpha \in X \Rightarrow \alpha' \in X) \wedge (\forall\alpha)(\alpha <_o \delta \wedge \mathrm{Lim}(\alpha) \wedge (\forall\beta)(\beta <_o \alpha \Rightarrow \beta \in X) \Rightarrow \alpha \in X)] \Rightarrow \delta \subseteq X$.

(c) (*Induction up to* ω.) $\vdash \emptyset \in X \wedge (\forall\alpha)(\alpha <_o \omega \wedge \alpha \in X \Rightarrow \alpha' \in X) \Rightarrow \omega \subseteq X$.

Proof

(a) Assume the antecedent. Let $Y = \{x \mid x \in On \wedge (\forall\alpha)(\alpha \leqslant_o x \Rightarrow \alpha \in X)\}$. It is easy to prove that $(\forall\alpha)(\alpha <_o \gamma \Rightarrow \alpha \in Y) \Rightarrow \gamma \in Y$. Hence, by Proposition 4.9, $On \subseteq Y$. But $Y \subseteq X$. Hence, $On \subseteq X$.

(b) The proof is left as an exercise.

(c) This is a special case of part (b), noting that $\vdash (\forall\alpha)(\alpha <_o \omega \Rightarrow \neg\mathrm{Lim}(\alpha))$.

Set theory depends heavily upon definitions by transfinite induction, which are justified by the following theorem.

PROPOSITION 4.14

(a) $\vdash (\forall X)(\exists_1 Y)(\mathrm{Fnc}(Y) \wedge \mathscr{D}(Y) = On \wedge (\forall\alpha)(Y`\alpha = X`(\alpha \restriction Y)))$. (Given X, there is a unique function Y defined on all ordinals such that the value of Y at α is the value of X applied to the restriction of Y to the set of ordinals less than α.)

(b) $\vdash (\forall x)(\forall X_1)(\forall X_2)(\exists_1 Y)(\mathrm{Fnc}(Y) \wedge \mathscr{D}(Y) = On \wedge Y`\emptyset = x \wedge (\forall\alpha)(Y`(\alpha') = X_1`(Y`\alpha)) \wedge (\forall\alpha)(\mathrm{Lim}(\alpha) \Rightarrow Y`\alpha = X_2`(\alpha \restriction Y)))$.

(c) (*Induction up to* δ.) $\vdash (\forall x)(\forall X_1)(\forall X_2)(\exists_1 Y)(\mathrm{Fnc}(Y) \wedge \mathscr{D}(Y) = \delta \wedge Y`\emptyset = x \wedge (\forall\alpha)(\alpha' <_o \delta \Rightarrow Y`(\alpha') = X_1`(Y`\alpha)) \wedge (\forall\alpha)(\mathrm{Lim}(\alpha) \wedge \alpha <_o \delta \Rightarrow Y`\alpha = X_2`(\alpha \restriction Y)))$.

Proof

(a) Let $Y_1 = \{u | \mathrm{Fnc}(u) \wedge \mathscr{D}(u) \in On \wedge (\forall\alpha)(\alpha \in \mathscr{D}(u) \Rightarrow u`\alpha = X`(\alpha \restriction u))\}$. Now, if $u_1 \in Y_1$ and $u_2 \in Y_1$, then $u_1 \subseteq u_2$ or $u_2 \subseteq u_1$. In fact, let $\gamma_1 = \mathscr{D}(u_1)$ and $\gamma_2 = \mathscr{D}(u_2)$. Either $\gamma_1 \leqslant_o \gamma_2$ or $\gamma_2 \leqslant_o \gamma_1$; say, $\gamma_1 \leqslant_o \gamma_2$. Let w be the set of ordinals $\alpha <_o \gamma_1$ such that $u_1`\alpha \neq u_2`\alpha$; assume $w \neq \emptyset$ and let η be the least ordinal in w. Then for all $\beta <_o \eta$, $u_1`\beta = u_2`\beta$. Hence, $u_1`\alpha = \eta \restriction u_2$. But $u_1`\eta = X`(\eta \restriction u_1)$ and $u_2`\eta = X`(\eta \restriction u_2)$; and so, $u_1`\eta = u_2`\eta$, contradicting our assumption. Therefore, $w = \emptyset$; that is, for all $\alpha \leqslant_o \gamma_1$, $u_1`\alpha = u_2`\alpha$. Hence, $u_1 = \gamma_1 \restriction u_1 = \gamma_1 \restriction u_2 \subseteq u_2$. Thus, any two functions in Y_1 agree in their common domain. Let $Y = \bigcup Y_1$. We leave it as an exercise to prove that Y is a function, the domain of which is either an ordinal or the class On, and $(\forall\alpha)(\alpha \in \mathscr{D}(Y) \Rightarrow Y`\alpha = X`(\alpha \restriction Y))$. That $\mathscr{D}(Y) = On$ follows easily from the observation that, if $\mathscr{D}(Y) = \delta$ and if we let $W = Y \cup \{\langle \delta, X`Y\rangle\}$, then $W \in Y_1$; so, $W \subseteq Y$ and $\delta \in \mathscr{D}(Y) = \delta$, which contradicts the fact that $\delta \notin \delta$. The uniqueness of Y follows by a simple transfinite induction (Proposition 4.9).

The proof of part (b) is similar to that of (a), and part (c) follows from (b).

Using Proposition 4.14, one can introduce new function letters by transfinite induction.

Examples

1. *Ordinal addition.* In Proposition 4.14(b), take

$$x = \beta \qquad X_1 = \{\langle u, v\rangle | v = u'\} \qquad X_2 = \{\langle u, v\rangle | v = \bigcup \mathscr{R}(u)\}$$

Hence, for each ordinal β, there is a unique function Y_β such that

$$Y_\beta{}'\emptyset = \beta \wedge (\forall\alpha)(Y_\beta{}'(\alpha') = (Y_\beta{}'\alpha)' \wedge [\mathrm{Lim}(\alpha) \Rightarrow Y_\beta{}'\alpha = \bigcup(Y_\beta{}''\alpha)])$$

Hence there is a unique binary function $+_o$ with domain $(On)^2$ such that, for any ordinals β and γ, $+_o(\beta,\gamma) = Y_\beta{}'\gamma$. As usual, we write $\beta +_o \gamma$ instead of $+_o(\beta,\gamma)$. Notice that:

$$\begin{aligned} \beta +_o \emptyset &= \beta \\ \beta +_o (\gamma') &= (\beta +_o \gamma)' \\ \mathrm{Lim}(\alpha) \Rightarrow \beta +_o \alpha &= \bigcup_{\tau <_o \alpha} (\beta +_o \tau) \end{aligned}$$

In particular,

$$\beta +_o 1 = \beta +_o (\emptyset') = (\beta +_o \emptyset)' = \beta'$$

2. *Ordinal multiplication.* In Proposition 4.14(b), take

$$x = \emptyset \quad X_1 = \{\langle u,v\rangle | v = u +_o \beta\} \qquad X_2 = \{\langle u,v\rangle | v = \bigcup \mathscr{R}(u)\}$$

Then, as in Example 1, one obtains a function $\beta \times_o \gamma$ with the properties

$$\begin{aligned} \beta \times_o \emptyset &= \emptyset \\ \beta \times_o (\gamma') &= (\beta \times_o \gamma) +_o \beta \\ \mathrm{Lim}(\alpha) \Rightarrow \beta \times_o \alpha &= \bigcup_{\tau <_o \alpha} (\beta \times_o \tau) \end{aligned}$$

Exercises

4.35 Prove: $\vdash \beta \times_o 1 = \beta \wedge \beta \times_o 2 = \beta +_o \beta$.

4.36 Justify the following definition of ordinal exponentiation.[†]

$$\begin{aligned} \exp(\beta,\emptyset) &= 1 \\ \exp(\beta,\gamma') &= \exp(\beta,\gamma) \times_o \beta \\ \mathrm{Lim}(\alpha) \Rightarrow \exp(\beta,\alpha) &= \bigcup_{\emptyset <_o \tau <_o \alpha} \exp(\beta,\tau) \end{aligned}$$

For any class X, let E_X be the membership relation restricted to X; that is, $E_X = \{\langle u,v\rangle | u \in v \wedge u \in X \wedge v \in X\}$.

[†]We use the notation $\exp(\beta,\alpha)$ instead of β^α in order to avoid confusion with the notation X^Y to be introduced later.

PROPOSITION 4.15[†]

Let R be a well-ordering relation on a class Y; that is, R We Y. Let F be a function from Y into Y such that, for any u and v in Y, if $\langle u, v\rangle \in R$, then $\langle F`u, F`v\rangle \in R$. Then, for all u in Y, $u = F`u$ or $\langle u, F`u\rangle \in R$.

Proof

Let $X = \{u | \langle F`u, u\rangle \in R\}$. We wish to show that $X = \emptyset$. Assume $X \neq \emptyset$. Since $X \subseteq Y$ and $R$ well-orders $Y$, there is an $R$-least element $u_0$ of $X$. Hence, $\langle F`u_0, u_0\rangle \in R$. Therefore $\langle F`(F`u_0), F`u_0\rangle \in R$. Thus, $F`u_0 \in X$, but $F`u_0$ is R-smaller than u_0, contradicting the definition of u_0.

COROLLARY 4.16

If Y is a class of ordinals, $F: Y \to Y$, and F is increasing on Y (that is, $\alpha \in Y \wedge \beta \in Y \wedge \alpha <_o \beta \Rightarrow F`\alpha <_o F`\beta$), then $\alpha \leqslant_o F`\alpha$ for all α in Y.

Proof

In Proposition 4.15, let R be E_Y. Note that E_Y well-orders Y, by Proposition 4.8(f) and Exercise 4.25.

COROLLARY 4.17

Let $\alpha <_o \beta$ and $y \subseteq \alpha$; that is, let y be a subset of a segment of β. Then $\langle E_\beta, \beta\rangle$ is not similar to $\langle E_y, y\rangle$.

Proof

Assume $\langle E_\beta, \beta\rangle$ is similar to $\langle E_y, y\rangle$. Then there is a function f from β onto y such that, for any u and v in $\beta, u <_o v \Leftrightarrow f`u <_o f`v$. Since the range of f is y, $f`\alpha \in y$. But $y \subseteq \alpha$. Hence $f`\alpha <_o \alpha$. But, by Corollary 4.16, $\alpha \leqslant_o f`\alpha$, which yields a contradiction.

[†]From this point on, we shall express many theorems of NBG in English by using the corresponding informal English translations. This is done to avoid writing lengthy wfs that are difficult to decipher and only in cases where the reader should be able to produce from the English version the precise wf of NBG.

COROLLARY 4.18

(a) For $\alpha \neq \beta$, $\langle E_\alpha, \alpha\rangle$ and $\langle E_\beta, \beta\rangle$ are not similar.
(b) For any α, if f is a similarity mapping of $\langle E_\alpha, \alpha\rangle$ with $\langle E_\alpha, \alpha\rangle$, then f is the identity mapping, that is, $f`\beta = \beta$ for all $\beta <_o \alpha$.

Proof

(a) Since $\alpha \neq \beta$, it follows by Proposition 4.8(d,c) that one of α and β is a segment of the other; say, α is a segment of β. Then Corollary 4.17 tells us that $\langle E_\beta, \beta\rangle$ is not similar to $\langle E_\alpha, \alpha\rangle$.
(b) By Corollary 4.16, $f`\beta \geqslant_o \beta$ for all $\beta <_o \alpha$. But, noting by Exercise 4.26(b) that $\breve{f}$ is a similarity mapping of $\langle E_\alpha, \alpha\rangle$ with $\langle E_\alpha, \alpha\rangle$, we again use Corollary 4.16 to conclude that $(\breve{f})`\beta \geqslant_o \beta$ for all $\beta <_o \alpha$. Hence $\beta = (\breve{f})`(f`\beta) \geqslant_o f`\beta \geqslant_o \beta$ and, therefore, $f`\beta = \beta$.

PROPOSITION 4.19

Assume that a non-empty set u is the field of a well-ordering r. Then there is a unique ordinal γ and a unique similarity mapping of $\langle E_\gamma, \gamma\rangle$ with $\langle r, u\rangle$.

Proof

Let $F = \{\langle v, w\rangle | w \in u - v \wedge (\forall z)(z \in u - v \Rightarrow \langle z, w\rangle \notin r)\}$. F is a function such that, if v is a subset of u and $u - v \neq \emptyset$, then $F`v$ is the $r$-least element of $u - v$. Let $X = \{\langle v, w\rangle | \langle \mathscr{R}(v), w\rangle \in F\}$. Now we use a definition by transfinite induction (Proposition 4.14) to obtain a function $Y$ with $On$ as its domain such that $(\forall\alpha)(Y`\alpha = X`(\alpha \restriction Y))$. Let $W = \{\alpha | Y``\alpha \subseteq u \wedge u - Y``\alpha \neq \emptyset\}$. Clearly, if $\alpha \in W$ and $\beta \in \alpha$, then $\beta \in W$. Hence, either $W = On$ or W is some ordinal γ. (If $W \neq On$, let γ be the least ordinal in $On - W$.) If $\alpha \in W$, then $Y`\alpha = X`(\alpha \restriction Y)$ is the r-least element of u-$Y``\alpha$; so, $Y`\alpha \in u$ and, if $\beta \in \alpha$, $Y`\alpha \neq Y`\beta$. Thus, Y is a one–one function on W and the range of Y restricted to W is a subset of u. Now, let $h = (W \restriction Y)$ and $f = \breve{h}$; that is, let f be the inverse of Y restricted to W. So, by the replacement axiom (R), W is a set. Hence, W is some ordinal γ. Let $g = \gamma \restriction Y$. Then g is a one–one function with domain γ and range a subset u_1 of u. We must show that $u_1 = u$ and that, if α and β are in γ and $\beta <_o \alpha$, then $\langle g`\beta, g`\alpha\rangle \in r$. Assume α and β are in γ and $\beta <_o \alpha$. Then $g``\beta \subseteq g``\alpha$ and, since $g`\alpha \in u - g``\alpha$, $g`\alpha \in u - g``\beta$. But $g`\beta$ is the r-least element of $u - g``\beta$. Hence, $\langle g`\beta, g`\alpha\rangle \in r$. It remains to prove that $u_1 = u$. Now, $u_1 = Y``\gamma$. Assume $u - u_1 \neq \emptyset$. Then $\gamma \in W$. But $W = \gamma$, which yields a contradiction. Hence, $u = u_1$. That γ is unique follows from Corollary 4.18(a).

Exercise

4.37 Show that the conclusion of Proposition 4.19 also holds when $u = \emptyset$ and that the unique ordinal γ is, in that case, $\emptyset$.

PROPOSITION 4.20

Let R be a well-ordering of a proper class X such that, for each $y \in X$, the class of all R-predecessors of y in X (i.e., the R-segment in X determined by y) is a set. Then R is 'similar' to E_{On}; that is, there is a (unique) one–one mapping H of On onto X such that $\alpha \in \beta \Leftrightarrow \langle H`\alpha, H`\beta\rangle \in R$.

Proof

Proceed as in the proof of Proposition 4.19. Here, however, $W = On$; also, one proves that $\mathscr{R}(Y) = X$ by using the hypothesis that every R-segment of X is a set. (If $X - \mathscr{R}(Y) \neq \emptyset$, then, if w is the R-least element of $X - \mathscr{R}(Y)$, the proper class On is the range of $\breve{Y}$, while the domain of $\breve{Y}$ is the R-segment of X determined by w, contradicting the replacement axiom.)

Exercise

4.38 Show that, if X is a proper class of ordinal numbers, then there is a unique one–one mapping H of On onto X such that $\alpha \in \beta \Leftrightarrow H`\alpha \in H`\beta$.

4.3 EQUINUMEROSITY. FINITE AND DENUMERABLE SETS

We say that two classes X and Y are *equinumerous* if and only if there is a one–one function F with domain X and range Y. We shall denote this by $X \cong Y$.

DEFINITIONS

$$X \underset{F}{\cong} Y \text{ for } \mathrm{Fnc}_1(F) \wedge \mathscr{D}(F) = X \wedge \mathscr{R}(F) = Y$$

$$X \cong Y \text{ for } (\exists F)(X \underset{F}{\cong} Y)$$

Notice that $\vdash (\forall x)(\forall y)(x \cong y \Leftrightarrow (\exists z)(x \underset{z}{\cong} y))$. Hence, a wf $x \cong y$ is predicative (that is, is equivalent to a wf using only set quantifiers).

Clearly, if $X \underset{F}{\cong} Y$, then $Y \underset{G}{\cong} X$, where $G = \breve{F}$. Also, if $X \underset{F_1}{\cong} Y$ and $Y \underset{F_2}{\cong} Z$ then $X \underset{H}{\cong} Z$, where H is the composition $F_2 \circ F_1$. Hence, we have the following result.

PROPOSITION 4.21

(a) $\vdash X \cong X$
(b) $\vdash X \cong Y \Rightarrow Y \cong X$
(c) $\vdash X \cong Y \wedge Y \cong Z \Rightarrow X \cong Z$

PROPOSITION 4.22

(a) $\vdash (X \cong Y \wedge Z \cong W \wedge X \cap Z = \emptyset \wedge Y \cap W = \emptyset) \Rightarrow X \cup Z \cong Y \cap W$
(b) $\vdash (X \cong Y \wedge Z \cong W) \Rightarrow X \times Z \cong Y \times W$
(c) $\vdash X \times \{y\} \cong X$
(d) $\vdash X \times Y \cong Y \times X$
(e) $\vdash (X \times Y) \times Z \cong X \times (Y \times Z)$

Proof

(a) Let $X \underset{F}{\cong} Y$ and $Z \underset{G}{\cong} W$. Then $X \cup Z \underset{H}{\cong} Y \cup W$, where $H = F \cup G$.

(b) Let $X \underset{F}{\cong} Y$ and $Z \underset{G}{\cong} W$. Let
$H = \{\langle u, v\rangle | (\exists x)(\exists y)(x \in X \wedge y \in Z \wedge u = \langle x, y\rangle \wedge v = \langle F`x, G`y\rangle)\}$
Then $X \times Z \underset{H}{\cong} Y \times W$.

(c) Let $F = \{\langle u, v\rangle | u \in X \wedge v = \langle u, y\rangle\}$. Then $X \underset{F}{\cong} X \times \{y\}$.

(d) Let $F = \{\langle u, v\rangle | (\exists x)(\exists y)(x \in X \wedge y \in Y \wedge u = \langle x, y\rangle \wedge v = (y, x\rangle)\}$.
Then $X \times Y \underset{F}{\cong} Y \times X$.

(e) Let $F = \{\langle u, v\rangle | (\exists x)(\exists y)(\exists z)(x \in X \wedge y \in Y \wedge z \in Z \wedge u = \langle\langle x, y\rangle, z\rangle \wedge v = \langle x, (y, z\rangle)\rangle)\}$. Then $(X \times Y) \times Z \underset{F}{\cong} X \times (Y \times Z)$.

DEFINITION

X^Y for $\{u | u: Y \to X\}$
X^Y is the class of all sets that are functions from Y into X.

Exercises

Prove the following.

4.39 $\vdash (\forall X)(\forall Y)(\exists X_1)(\exists Y_1)(X \cong X_1 \wedge Y \cong Y_1 \wedge X_1 \cap Y_1 = \emptyset)$
4.40 $\vdash \mathscr{P}(y) \cong 2^y$ (Recall that $2 = \{\emptyset, 1\}$ and $1 = \{\emptyset\}$.)
4.41 (a) $\vdash \neg \mathrm{M}(Y) \Rightarrow X^Y = \emptyset$
(b) $\vdash (\forall x)(\forall y)\mathrm{M}(x^y)$
4.42 (a) $\vdash X^\emptyset = 1$
(b) $\vdash 1^y \cong 1$
(c) $\vdash Y \neq \emptyset \Rightarrow \emptyset^Y = \emptyset$
4.43 $\vdash X \cong X^{\{u\}}$
4.44 $\vdash X \cong Y \wedge Z \cong W \Rightarrow X^Z \cong Y^W$
4.45 $\vdash X \cap Y = \emptyset \Rightarrow Z^{X \cup Y} \cong Z^X \times Z^Y$
4.46 $\vdash (\forall x)(\forall y)(\forall z)[(x^y)^z \cong x^{y \times z}]$
4.47 $\vdash (X \times Y)^Z \cong X^Z \times Y^Z$
4.48 $\vdash (\forall x)(\forall R)(R \text{ We } x \Rightarrow (\exists \alpha)(x \cong \alpha))$

We can define a partial order $\preccurlyeq$ on classes such that, intuitively, $X \preccurlyeq Y$ if and only if Y has at least as many elements as X.

DEFINITIONS

$X \preccurlyeq Y$ for $(\exists Z)(Z \subseteq Y \wedge X \cong Z)$
(X is equinumerous with a subclass of Y)
$X \prec Y$ for $X \preccurlyeq Y \wedge \neg(X \cong Y)$
(Y is strictly greater in size than X)

Exercises

Prove the following.

4.49 $\vdash X \preccurlyeq Y \Leftrightarrow (X \prec Y \vee X \cong Y)$
4.50 $\vdash X \preccurlyeq Y \wedge \neg \mathrm{M}(X) \Rightarrow \neg \mathrm{M}(Y)$
4.51 $\vdash X \preccurlyeq Y \wedge (\exists Z)(Z \text{ We } Y) \Rightarrow (\exists Z)(Z \text{ We } X)$
4.52 $\vdash (\forall \alpha)(\forall \beta)(\alpha \preccurlyeq \beta \vee \beta \preccurlyeq \alpha)$ [*Hint*: Proposition 4.8(k).]

PROPOSITION 4.23

(a) $\vdash X \preccurlyeq X \wedge \neg(X \prec X)$
(b) $\vdash X \subseteq Y \Rightarrow X \preccurlyeq Y$
(c) $\vdash X \preccurlyeq Y \wedge Y \preccurlyeq Z \Rightarrow X \preccurlyeq Z$
(d) $\vdash X \preccurlyeq Y \wedge Y \preccurlyeq X \Rightarrow X \cong Y$ (Bernstein's theorem)

Proof

(a), (b) These proofs are obvious.

(c) Assume $X \underset{F}{\cong} Y_1 \wedge Y_1 \subseteq Y \wedge Y \underset{G}{\cong} Z_1 \wedge Z_1 \subseteq Z$. Let H be the composition of F and G. Then $\mathscr{R}(H) \subseteq Z \wedge X \underset{H}{\cong} \mathscr{R}(H)$. So, $X \preccurlyeq Z$.

(d) There are many proofs of this nontrivial theorem. The following one was devised by Hellman (1961). First we derive a lemma.

Lemma. Assume $X \cap Y = \emptyset$, $X \cap Z = \emptyset$ and $Y \cap Z = \emptyset$, and let $X \underset{F}{\cong} X \cup Y \cup Z$. Then there is a G such that $X \underset{G}{\cong} X \cup Y$.

Proof. Define a function H on a subclass of $X \times \omega$ as follows: $\langle\langle u, k\rangle, v\rangle \in H$ if and only if $u \in X$ and $k \in \omega$ and there is a function f with domain k' such that $f`\emptyset = F`u$ and, if $j \in k$, then $f`j \in X$ and $f`(j') = F`(f`j)$ and $f`k = v$. Thus, $H`(\langle u, \emptyset\rangle) = F`u, H`(\langle u, 1\rangle) = F`(F`u)$ if $F`u \in X$, and $H`(\langle u, 2\rangle) = F`(F`(F`u))$ if $F`u$ and $F`(F`u)$ are in X, and so on. Let X^* be the class of all u in X such that $(\exists y)(y \in \omega \wedge \langle u, y\rangle \in \mathscr{D}(H) \wedge H`(\langle u, y\rangle) \in Z)$. Let $Y^*$ be the class of all $u$ in $X$ such that $(\forall y)(y \in \omega \wedge \langle u, y\rangle \in \mathscr{D}(H) \Rightarrow H`(\langle u, y\rangle) \notin Z)$. Then $X = X^* \cup Y^*$. Now define G as follows: $\mathscr{D}(G) = X$ and, if $u \in X^*$, then $G`u = u$, whereas, if $u \in Y^*$, then $G`u = F`u$. Then $X \underset{G}{\cong} X \cup Y$. (This is left as an exercise.)

Now, to prove Bernstein's theorem, assume $X \underset{F}{\cong} Y_1 \wedge Y_1 \subseteq Y \wedge Y \underset{G}{\cong} X_1 \wedge X_1 \subseteq X$. Let $A = G``Y_1 \subseteq X_1 \subseteq X$. But $A \cap (X_1 - A) = \emptyset, A \cap (X - X_1) = \emptyset$ and $(X - X_1) \cap (X_1 - A) = \emptyset$. Also, $X = (X - X_1) \cup (X_1 - A) \cup A$, and the compositon H of F and G is a one–one function with domain X and range A. Hence, $A \underset{H}{\cong} X$. So, by the lemma, there is a one–one function D such that $A \underset{D}{\cong} X_1$ (since $(X_1 - A) \cup A = X_1$). Let T be the composition of the functions H, D and $\breve{G}$; that is, $T`u = (\breve{G})`(D`(H`u))$. Then $X \underset{T}{\cong} Y$, since $X \underset{H}{\cong} A$ and $A \underset{D}{\cong} X_1$ and $X_1 \underset{G}{\cong} Y$.

Exercises

4.53 Carry out the details of the following proof (due to J. Whitaker) of Bernstein's theorem in the case where X and Y are sets. Let $X \underset{F}{\cong} Y_1 \wedge Y_1 \subseteq Y \wedge Y \underset{G}{\cong} X_1 \wedge X_1 \subseteq X$. We wish to find a set $Z \subseteq X$ such that G, restricted to $Y - F``Z$, is a one–one function of $Y - F``Z$ onto $X - Z$. [If we have such a set Z, let $H = (Z \restriction F) \cup ((X - Z) \restriction G)$; that is, $H`x = F`x$ for $x \in Z$, and $H`x = \breve{G}`x$ for $x \in X - Z$. Then $X \underset{H}{\cong} Y$.] Let $Z = \{x \mid (\exists u)(u \subseteq X \wedge x \in u \wedge G``(Y - F``u) \subseteq X - u)\}$. Notice that this proof does not presuppose the definition of ω nor any other part of the theory of ordinals.

4.54 Prove: (a) $\vdash X \preccurlyeq X \cup Y$ (b) $\vdash X \prec Y \Rightarrow \neg(Y \prec X)$ (c) $\vdash X \prec Y \wedge Y \preccurlyeq Z \Rightarrow X \prec Z$

PROPOSITION 4.24 Assume $X \preccurlyeq Y$ and $A \preccurlyeq B$. Then:

(a) $Y \cap B = \emptyset \Rightarrow X \cup A \preccurlyeq Y \cup B$
(b) $X \times A \preccurlyeq Y \times B$
(c) $X^A \preccurlyeq Y^B$ if B is a set and $\neg(X = A = Y = \emptyset \wedge B \neq \emptyset)$

Proof

(a) Assume $X \underset{F}{\cong} Y_1 \subseteq Y$ and $A \underset{G}{\cong} B_1 \subseteq B$. Let H be a function with domain $X \cup A$ such that $H`x = F`x$ for $x \in X$, and $H`x = G`x$ for $x \in A - X$. Then $X \cup A \underset{H}{\cong} H``(X \cup A) \subseteq Y \cup B$.
(b) and (c) are left as exercises.

PROPOSITION 4.25

(a) $\vdash \neg(\exists f)(\mathrm{Fnc}(f) \wedge \mathscr{D}(f) = x \wedge \mathscr{R}(f) = \mathscr{P}(x))$. (There is no function from x onto $\mathscr{P}(x)$.)
(b) $\vdash x \prec \mathscr{P}(x)$ (Cantor's theorem)

Proof

(a) Assume $\mathrm{Fnc}(f) \wedge \mathscr{D}(f) = x \wedge \mathscr{R}(f) = \mathscr{P}(x)$. Let $y = \{u | u \in x \wedge u \notin f`u\}$. Then $y \in \mathscr{P}(x)$. Hence, there is some $z$ in $x$ such that $f`z = y$. But, $(\forall u)(u \in y \Leftrightarrow u \in x \wedge u \notin f`u)$. Hence, $(\forall u)(u \in f`z \Leftrightarrow u \in x \wedge u \notin f`u)$. By rule A4, $z \in f`z \Leftrightarrow z \in x \wedge z \notin f`z$. Since $z \in x$, we obtain $z \in f`z \Leftrightarrow z \notin f`z$, which yields a contradicition.
(b) Let f be the function with domain x such that $f`u = \{u\}$ for each u in x. Then $f``x \subseteq \mathscr{P}(x)$ and f is one–one. Hence, $x \preccurlyeq \mathscr{P}(x)$. By part (a), $x \cong \mathscr{P}(x)$ is impossible. Hence, $x \prec \mathscr{P}(x)$.

In naive set theory, Proposition 4.25(b) gives rise to Cantor's paradox. If we let $x = V$, then $V \prec \mathscr{P}(V)$. But $\mathscr{P}(V) \subseteq V$ and, therefore, $\mathscr{P}(V) \preccurlyeq V$. From $V \prec \mathscr{P}(V)$, we have $V \preccurlyeq \mathscr{P}(V)$. By Bernstein's theorem, $V \cong \mathscr{P}(V)$, contradicting $V \prec \mathscr{P}(V)$. In NBG, this argument is just another proof that V is not a set.

Notice that we have not proved $\vdash (\forall x)(\forall y)(x \preccurlyeq y \vee y \preccurlyeq x)$. This intuitively plausible statement is, in fact, not provable, since it turns out to be equivalent to the axiom of choice (which will be discussed in Section 4.5).

The equinumerosity relation $\cong$ has all the properties of an equivalence relation. We are inclined, therefore, to partition the class of all sets into equivalence classes under this relation. The equivalence class of a set x would be the class of all sets equinumerous with x. The equivalence classes are called *Frege–Russell cardinal numbers*. For example, if u is a set and

$x = \{u\}$, then the equivalence class of x is the class of all singletons $\{v\}$ and is referred to as the cardinal number 1_c. Likewise, if $u \neq v$ and $y = \{u, v\}$, then the equivalence class of y is the class of all sets that contain exactly two elements and would be the cardinal number 2_c; that is 2_c is $\{x|(\exists w)(\exists z)(w \neq z \wedge x = \{w, z\})\}$. All the Frege–Russell cardinal numbers, except the cardinal number 0_c of $\emptyset$ (which is $\{\emptyset\}$), turn out to be proper classes. For example, $V \cong 1_c$. (Let $F'x = \{x\}$ for all x. Then $V \underset{F}{\cong} 1_c$.) But, $\neg M(V)$. Hence, by the replacement axiom, $\neg M(1_c)$.

Exercise

4.55 Prove $\vdash \neg M(2_c)$.

Because all the Frege–Russell cardinal numbers (except 0_c) are proper classes, we cannot talk about classes of such cardinal numbers, and it is difficult or impossible to say and prove many interesting things about them. Most assertions one would like to make about cardinal numbers can be paraphrased by the suitable use of $\cong$, $\preccurlyeq$ and $\prec$. However, we shall see later that, given certain additional plausible axioms, there are other ways of defining a notion that does essentially the same job as the Frege–Russell cardinal numbers.

To see how everything we want to say about cardinal numbers can be said without explicit mention of cardinal numbers, consider the following treatment of the 'sum' of cardinal numbers.

DEFINITION

$X +_c Y$ for $(X \times \{\emptyset\}) \cup (Y \times \{1\})$

Note that $\vdash \emptyset \neq 1$ (since 1 is $\{\emptyset\}$). Hence, $X \times \{\emptyset\}$ and $Y \times \{1\}$ are disjoint and, therefore, their union is a class whose 'size' is the sum of the 'sizes' of X and Y.

Exercise

4.56 Prove:

(a) $\vdash X \preccurlyeq X +_c Y \wedge Y \preccurlyeq X +_c Y$

(b) $\vdash X \cong A \wedge Y \cong B \Rightarrow X +_c Y \cong A +_c B$

(c) $\vdash X +_c Y \cong Y +_c X$

(d) $\vdash M(X +_c Y) \Leftrightarrow M(X) \wedge M(Y)$

(e) $\vdash X +_c (Y +_c Z) \cong (X +_c Y) +_c Z$

(f) $\vdash X \preccurlyeq Y \Rightarrow X +_c Z \preccurlyeq Y +_c Z$

(g) $\vdash X +_c X = X \times 2$ (Recall that 2 is $\{\emptyset, 1\}$.)

(h) $\vdash X^{Y +_c Z} \cong X^Y \times X^Z$

(i) $\vdash x \cong x +_c 1 \Rightarrow 2^x +_c x \cong 2^x$

Finite sets

Remember that ω is the set of all ordinals α such that α and all smaller ordinals are successor ordinals or $\emptyset$. The elements of ω are called *finite ordinals*, and the elements of $On - \omega$ are called *infinite ordinals*. From an intuitive standpoint, ω consists of $\emptyset, 1, 2, 3, \ldots$, where each term in this sequence after $\emptyset$ is the successor of the preceding term. Note that $\emptyset$ contains no members, $1 = \{\emptyset\}$ and contains one member, $2 = \{\emptyset, 1\}$ and contains two members, $3 = \{\emptyset, 1, 2\}$ and contains three members, etc. Thus, it is reasonable to think that, for each intuitive finite number n, there is exactly one finite ordinal that contains exactly n members. So, if a class has n members, it should be equinumerous with a finite ordinal. Therefore, a class will be called *finite* if and only if it is equinumerous with a finite ordinal.

DEFINITION

$$\mathrm{Fin}(X) \text{ for } (\exists\alpha)(\alpha \in \omega \wedge X \cong \alpha) \qquad (X \text{ is finite})$$

Exercise

4.57 Prove:

(a) $\vdash \mathrm{Fin}(X) \Rightarrow \mathrm{M}(X)$ (Every finite class is a set)
(b) $\vdash (\forall\alpha)(\alpha \in \omega \Rightarrow \mathrm{Fin}(\alpha))$ (Every finite ordinal is finite.)
(c) $\vdash \mathrm{Fin}(X) \wedge X \cong Y \Rightarrow \mathrm{Fin}(Y)$

PROPOSITION 4.26

(a) $\vdash (\forall\alpha)(\alpha \notin \omega \Rightarrow \alpha \cong \alpha')$.
(b) $\vdash (\forall\alpha)(\forall\beta)(\alpha \in \omega \wedge \alpha \neq \beta \Rightarrow \neg(\alpha \cong \beta))$. (No finite ordinal is equinumerous with any other ordinal.)
(c) $\vdash (\forall\alpha)(\forall x)(\alpha \in \omega \wedge x \subset \alpha \Rightarrow \neg(\alpha \cong x))$. (No finite ordinal is equinumerous with a proper subset of itself.)

Proof

(a) Assume $\alpha \notin \omega$. Define a function f with domain α' as follows: $f`\delta = \delta'$ if $\delta \in \omega$; $f`\delta = \delta$ if $\delta \in \alpha' \wedge \delta \notin \omega \cup \{\alpha\}$; and $f`\alpha = \emptyset$. Then $\alpha' \underset{f}{\cong} \alpha$.
(b) Assume this is false, and let α be the least ordinal such that $\alpha \in \omega$ and there is $\beta \neq \alpha$ such that $\alpha \cong \beta$. Hence, $\alpha <_o \beta$. (Otherwise, β would be a smaller ordinal than α and β would also be in ω, and β would be equinumerous with another ordinal, namely, α.) Let $\alpha \underset{f}{\cong} \beta$. If $\alpha = \emptyset$,

then $f = \emptyset$ and $\beta = \emptyset$, contradicting $\alpha \neq \beta$. So, $\alpha \neq \emptyset$. Since $\alpha \in \omega, \alpha = \delta'$ for some $\delta \in \omega$. We may assume that $\beta = \gamma'$ for some γ. (If $\beta \in \omega$, then $\beta \neq \emptyset$; and if $\beta \notin \omega$, then, by part (a), $\beta \cong \beta'$ and we can take β' instead of β.) Thus, $\delta' = \alpha \underset{f}{\cong} \gamma'$. Also, $\delta \neq \gamma$, since $\alpha \neq \beta$.

Case 1. $f`\delta = \gamma$. Then $\delta \underset{g}{\cong} \gamma$, where $g = \delta \restriction f$.

Case 2. $f`\delta \neq \gamma$. Then there is some $\mu \in \delta$ such that $f`\mu = \gamma$. Let $h = ((\delta \restriction f) - \{\langle \mu, \gamma \rangle\}) \cup \{\langle \mu, f`\delta \rangle\}$; that is, let $h`\tau = f`\tau$ if $\tau \notin \{\delta, \mu\}$, and $h`\mu = f`\delta$. Then $\delta \underset{h}{\cong} \gamma$.

In both cases, δ is a finite ordinal smaller than α that is equinumerous with a different ordinal γ, contradicting the minimality of α.

(c) Assume $\beta \in \omega \wedge x \subset \beta \wedge \beta \cong x$ holds for some β, and let α be the least such β. Clearly, $\alpha \neq \emptyset$; hence, $\alpha = \gamma'$ for some γ. But, as in the proof of part (b), one can then show that γ is also equinumerous with a proper subset of itself, contradicting the minimality of α.

Exercises

4.58 Prove: $\vdash (\forall \alpha)(\mathrm{Fin}(\alpha) \Leftrightarrow \alpha \in \omega)$.

4.59 Prove that the axiom of infinity (I) is equivalent to the following sentence.

$$(*) \qquad (\exists x)((\exists u)(u \in x) \wedge (\forall y)(y \in x \Rightarrow (\exists z)(z \in x \wedge y \subset z)))$$

PROPOSITION 4.27

(a) $\vdash \mathrm{Fin}(X) \wedge Y \subseteq X \Rightarrow \mathrm{Fin}(Y)$
(b) $\vdash \mathrm{Fin}(X) \Rightarrow \mathrm{Fin}(X \cup \{y\})$
(c) $\vdash \mathrm{Fin}(X) \wedge \mathrm{Fin}(Y) \Rightarrow \mathrm{Fin}(X \cup Y)$

Proof

(a) Assume $\mathrm{Fin}(X) \wedge Y \subseteq X$. Then $X \cong \alpha$, where $\alpha \in \omega$. Let $g = Y \restriction f$ and $W = g``Y \subseteq \alpha$. W is a set of ordinals, and so, E_W is a well-ordering of W. By Proposition 4.19, $\langle E_W, W \rangle$ is similar to $\langle E_\beta, \beta \rangle$ for some ordinal β. Hence, $W \cong \beta$. In addition, $\beta \leqslant_o \alpha$. (If $\alpha <_o \beta$, then the similarity of $\langle E_\beta, \beta \rangle$ to $\langle E_W, W \rangle$ contradicts Corollary 4.17.) Since $\alpha \in \omega, \beta \in \omega$. From $Y \underset{g}{\cong} W \wedge W \cong \beta$, it follows that $\mathrm{Fin}(Y)$.

(b) If $y \in X$, then $X \cup \{y\} = X$ and the result is trivial. So, assume $y \notin X$. From $\mathrm{Fin}(X)$ it follows that there is a finite ordinal α and a function f such that $\alpha \underset{f}{\cong} X$. Let $g = f \cup \{\langle \alpha, y \rangle\}$. Then $\alpha' \underset{g}{\cong} X \cup \{y\}$. Hence, $\mathrm{Fin}(X \cup \{y\})$.

(c) Let $Z = \{u | u \in \omega \wedge (\forall x)(\forall y)(\forall f)(x \underset{f}{\cong} u \wedge \text{Fin}(y) \Rightarrow \text{Fin}(x \cup y))\}$. We must show that $Z = \omega$. Clearly, $\emptyset \in Z$, for if $x \cong \emptyset$, then $x = \emptyset$ and $x \cup y = y$. Assume that $\alpha \in Z$. Let $x \underset{f}{\cong} \alpha'$ and $\text{Fin}(y)$. Let w be such that $f`w = \alpha$ and let $x_1 = x - \{w\}$. Then $x_1 \cong \alpha$. Since $\alpha \in Z$, $\text{Fin}(x_1 \cup y)$. But $x \cup y = (x_1 \cup y) \cup \{w\}$. Hence, by part (b), $\text{Fin}(x \cup y)$. Thus, $\alpha' \in Z$. Hence, by Proposition 4.11(c), $Z = \omega$.

DEFINITIONS

$\text{DedFin}(X)$ for $\text{M}(X) \wedge (\forall Y)(Y \subset X \Rightarrow \neg(X \cong Y))$
(X is *Dedekind-finite*, that is, X is a set that is not equinumerous with any proper subset of itself)

$\text{DedInf}(X)$ for $\text{M}(X) \wedge \neg\text{DedFin}(X)$
(X is *Dedekind-infinite*, that is, X is a set that is equinumerous with a proper subset of itself)

COROLLARY 4.28

$(\forall x)(\text{Fin}(x) \Rightarrow \text{DedFin}(x))$ (Every finite set is Dedekind-finite)†

Proof

This follows easily from Proposition 4.26(c) and the definition of 'finite'.

DEFINITIONS

$\text{Inf}(X)$ for $\neg\text{Fin}(X)$	(X is *infinite*)
$\text{Den}(X)$ for $X \cong \omega$	(X is *denumerable*)
$\text{Count}(X)$ for $\text{Fin}(X) \vee \text{Den}(X)$	(X is *countable*)

Exercise

4.60 Prove:

(a) $\vdash \text{Inf}(X) \wedge X \cong Y \Rightarrow \text{Inf}(Y)$
(b) $\vdash \text{Den}(X) \wedge X \cong Y \Rightarrow \text{Den}(Y)$
(c) $\vdash \text{Den}(X) \Rightarrow \text{M}(X)$
(d) $\vdash \text{Count}(X) \wedge X \cong Y \Rightarrow \text{Count}(Y)$
(e) $\vdash \text{Count}(X) \Rightarrow \text{M}(X)$

†The converse is not provable without additional assumptions, such as the axiom of choice.

PROPOSITION 4.29

(a) $\vdash \mathrm{Inf}(X) \wedge X \subseteq Y \Rightarrow \mathrm{Inf}(Y)$
(b) $\vdash \mathrm{Inf}(X) \Leftrightarrow \mathrm{Inf}(X \cup \{y\})$
(c) $\vdash \mathrm{DedInf}(X) \Rightarrow \mathrm{Inf}(X)$
(d) $\vdash \mathrm{Inf}(\omega)$

Proof

(a) This follows from Proposition 4.27(a).
(b) $\vdash \mathrm{Inf}(X) \Rightarrow \mathrm{Inf}(X \cup \{y\})$ by part (a), and $\vdash \mathrm{Inf}(X \cup \{y\}) \Rightarrow \mathrm{Inf}(X)$ by Proposition 4.27(b)
(c) Use Corollary 4.28.
(d) $\vdash \omega \notin \omega$. If $\mathrm{Fin}(\omega)$, then $\omega \cong \alpha$ for some α in ω, contradicting Proposition 4.26(b).

PROPOSITION 4.30

$\vdash (\forall v)(\forall z)(\mathrm{Den}(v) \wedge z \subseteq v \Rightarrow \mathrm{Count}(z))$. (Every subset of a denumerable set is countable.)

Proof

It suffices to prove that $z \subseteq \omega \Rightarrow \mathrm{Fin}(z) \vee \mathrm{Den}(z)$. Assume $z \subseteq \omega \wedge \neg\mathrm{Fin}(z)$. Since $\neg\mathrm{Fin}(z)$, for any α in z, there is some β in z with $\alpha <_0 \beta$. (Otherwise, $z \subseteq \alpha'$ and, since $\mathrm{Fin}(\alpha')$, $\mathrm{Fin}(z)$.), Let X be a function such that, for any α in ω, $X`\alpha$ is the least ordinal $\beta$ in $z$ with $\alpha <_0 \beta$. Then, by Proposition 4.14(c) (with $\delta = \omega$), there is a function $Y$ with domain $\omega$ such that $Y`\emptyset$ is the least ordinal in z and, for any γ in ω, $Y`(\gamma')$ is the least ordinal $\beta$ in $z$ with $\beta >_0 Y`\gamma$. Clearly, Y is one–one, $\mathscr{D}(Y) = \omega$, and $Y``\omega \subseteq z$. To show that $\mathrm{Den}(z)$, it suffices to show that $Y``\omega = z$. Assume $z - Y``\omega \neq \emptyset$. Let $\delta$ be the least ordinal in $z - Y``\omega$, and let τ be the least ordinal in $Y``\omega$ with $\tau >_0 \delta$. Then $\tau = Y`\sigma$ for some $\sigma$ in $\omega$. Since $\delta <_0 \tau, \sigma \neq \emptyset$. So, $\sigma = \mu'$ for some $\mu$ in $\omega$. Then $\tau = Y`\sigma$ is the least ordinal in z that is greater than $Y`\mu$. But $\delta >_0 Y`\mu$, since τ is the least ordinal in $Y``\omega$ that is greater than δ. Hence, $\tau \leqslant_0 \delta$, which contradicts $\delta <_0 \tau$.

Exercises

4.61 Prove: $\vdash \mathrm{Count}(X) \wedge Y \subseteq X \Rightarrow \mathrm{Count}(Y)$.
4.62 Prove:
(a) $\vdash \mathrm{Fin}(X) \Rightarrow \mathrm{Fin}(\mathscr{P}(X))$
(b) $\vdash \mathrm{Fin}(X) \wedge (\forall y)(y \in X \Rightarrow \mathrm{Fin}(y)) \Rightarrow \mathrm{Fin}(\bigcup x)$

(c) $\vdash X \preccurlyeq Y \wedge \mathrm{Fin}(Y) \Rightarrow \mathrm{Fin}(X)$
(d) $\vdash \mathrm{Fin}(\mathscr{P}(X)) \Rightarrow \mathrm{Fin}(X)$
(e) $\vdash \mathrm{Fin}(\bigcup X) \Rightarrow \mathrm{Fin}(X) \wedge (\forall y)(y \in X \Rightarrow \mathrm{Fin}(y))$
(f) $\vdash \mathrm{Fin}(X) \Rightarrow (X \preccurlyeq Y \vee Y \preccurlyeq X)$
(g) $\vdash \mathrm{Fin}(X) \wedge \mathrm{Inf}(Y) \Rightarrow X \prec Y$
(h) $\vdash \mathrm{Fin}(X) \wedge Y \subset X \Rightarrow Y \prec X$
(i) $\vdash \mathrm{Fin}(X) \wedge \mathrm{Fin}(Y) \Rightarrow \mathrm{Fin}(X \times Y)$
(j) $\vdash \mathrm{Fin}(X) \wedge \mathrm{Fin}(Y) \Rightarrow \mathrm{Fin}(X^Y)$
(k) $\vdash \mathrm{Fin}(X) \wedge y \notin X \Rightarrow X \prec X \cup \{y\}$

4.63 Define X to be a *minimal* (respectively, *maximal*) element of Y if and only if $X \in Y$ and $(\forall y)(y \in Y \Rightarrow \neg(y \subset X))$(respectively, $(\forall y)(y \in Y \Rightarrow \neg (X \subset y))$). Prove that a set Z is finite if and only if every non-empty set of subsets of Z has a minimal (respectively, maximal) element (Tarski, 1925).

4.64 Prove:

(a) $\vdash \mathrm{Fin}(X) \wedge \mathrm{Den}(Y) \Rightarrow \mathrm{Den}(X \cup Y)$
(b) $\vdash \mathrm{Fin}(X) \wedge \mathrm{Den}(Y) \wedge X \neq \emptyset \Rightarrow \mathrm{Den}(X \times Y)$
(c) $\vdash (\forall x)[\mathrm{DedInf}(x) \Leftrightarrow (\exists y)(y \subseteq x \wedge \mathrm{Den}(y))]$. (A set is Dedekind-infinite if and only if it has a denumerable subset)
(d) $\vdash (\forall x)[(\exists y)(y \subseteq x \wedge \mathrm{Den}(y)) \Leftrightarrow \omega \preccurlyeq x]$
(e) $\vdash (\forall \alpha)[(\alpha \notin \omega \Rightarrow \mathrm{DedInf}(\alpha)) \wedge (\forall \alpha)(\mathrm{Inf}(\alpha) \Rightarrow \alpha \notin \omega)$
(f) $\vdash (\forall x)(\forall y)(y \notin x \Rightarrow [\mathrm{DedInf}(x) \Leftrightarrow x \cong x \cup \{y\}])$
(g) $\vdash (\forall x)(\omega \preccurlyeq x \Leftrightarrow x +_c 1 \cong x)$

4.65 If NBG is consistent, then, by Proposition 2.17, NBG has a denumerable model. Explain why this does not contradict Cantor's theorem, which implies that there exist non-denumerable infinite sets (such as $\mathscr{P}(\omega)$). This apparent, but not genuine, contradiction is sometimes called *Skolem's paradox*.

4.4 HARTOGS' THEOREM. INITIAL ORDINALS. ORDINAL ARITHMETIC

An unjustly neglected proposition with many uses in set theory is Hartogs' theorem.

PROPOSITION 4.31 (HARTOGS, 1915)

$\vdash (\forall x)(\exists \alpha)(\forall y)(y \subseteq x \Rightarrow \neg(\alpha \cong y))$. (For any set x, there is an ordinal that is not equinumerous with any subset of x.)

Proof

Assume that every ordinal α is equinumerous with some subset y of x. Hence, $y \overset{f}{\cong} \alpha$ for some f. Define a relation r on y by stipulating that $\langle u, v\rangle \in r$ if and only if $f`u \in f`v$. Then r is a well-ordering of y such that $\langle r, y\rangle$ is similar to $\langle \mathrm{E}_\alpha, \alpha\rangle$. Now define a function F with domain On such that, for any α, $F`\alpha$ is the set w of all pairs $\langle z, y\rangle$ such that $y \subseteq x$, z is a well-ordering of y, and $\langle \mathrm{E}_\alpha, \alpha\rangle$ is similar to $\langle z, y\rangle$. (w is a set, since $w \subseteq \mathscr{P}(x \times x) \times \mathscr{P}(x)$.) Since, $F``(On) \subseteq \mathscr{P}(\mathscr{P}(x \times x) \times \mathscr{P}(x))$, $F``(On)$ is a set. F is one–one; hence, $On = \breve{F}``(F``(On))$ is a set by the replacement axiom, contradicting Proposition 4.8(h).

DEFINITION

Let $\mathscr{H}$ denote the function with domain V such that, for every x, $\mathscr{H}`x$ is the least ordinal α that is not equinumerous with any subset of x. ($\mathscr{H}$ is called *Hartogs' function.*)

COROLLARY 4.32

$$(\forall x)(\mathscr{H}`x \preccurlyeq \mathscr{P}\mathscr{P}\mathscr{P}\mathscr{P}(x))$$

Proof

With each $\beta <_o \mathscr{H}`x$, associate the set of relations $r$ such that $r \subseteq x \times x$, $r$ is a well-ordering of its field $y$, and $\langle r, y\rangle$ is similar to $\langle \mathrm{E}_\beta, \beta\rangle$. This defines a one–one function from $\mathscr{H}`x$ into $\mathscr{P}\mathscr{P}(x \times x)$. Hence, $\mathscr{H}`x \preccurlyeq \mathscr{P}\mathscr{P}(x \times x)$. By Exercise 4.12(s), $x \times x \subseteq \mathscr{P}\mathscr{P}(x)$. So, $\mathscr{P}\mathscr{P}(x \times x) \subseteq \mathscr{P}\mathscr{P}\mathscr{P}\mathscr{P}(x)$, and therefore, $\mathscr{H}`x \preccurlyeq \mathscr{P}\mathscr{P}\mathscr{P}\mathscr{P}(x)$.

DEFINITION

$$\mathrm{Init}(X) \text{ for } X \in On \wedge (\forall \beta)(\beta <_o X \Rightarrow \neg(\beta \cong X))$$
(X is an *initial ordinal*)

An initial ordinal is an ordinal that is not equinumerous with any smaller ordinal.

Exercises

4.66 (a) $\vdash (\forall \alpha)(\alpha \in \omega \Rightarrow \mathrm{Init}(\alpha))$. (Every finite ordinal is an initial ordinal.)
(b) $\vdash \mathrm{Init}(\omega)$. [*Hint*: Use Proposition 4.26(b) for both parts.]

4.67 Prove:
(a) For every x, $\mathscr{H}'x$ is an initial ordinal.
(b) For any ordinal α, $\mathscr{H}'\alpha$ is the least initial ordinal greater than α.
(c) For any set x, $\mathscr{H}'x = \omega$ if any only if x is infinite and x is Dedekind-finite. [*Hint*: Exercise 4.64(c).]

Definition by transfinite induction (Proposition 4.14(b)) yields a function G with domain On such that

$$G'\emptyset = \omega$$
$$G'(\alpha') = \mathscr{H}'(G'\alpha) \quad \text{for every } \alpha$$
$$G'\lambda = \bigcup(G''(\lambda)) \quad \text{for every limit ordinal } \lambda$$

PROPOSITION 4.33

(a) $\vdash (\forall\alpha)(\mathrm{Init}(G'\alpha) \wedge \omega \leqslant_0 G'\alpha \wedge (\forall\beta)(\beta <_0 \alpha \Rightarrow G'\beta <_0 G'\alpha))$
(b) $\vdash (\forall\alpha)(\alpha \leqslant_0 G'\alpha)$
(c) $\vdash (\forall\beta)(\omega \leqslant_0 \beta \wedge \mathrm{Init}(\beta) \Rightarrow (\exists\alpha)(G'\alpha = \beta))$

Proof

(a) Let $X = \{\alpha \mid \mathrm{Init}(G'\alpha) \wedge \omega \leqslant_0 G'\alpha \wedge (\forall\beta)(\beta <_0 \alpha \Rightarrow G'\beta <_0 G'\alpha)\}$.

We must show that $On \subseteq X$. To do this, we use the second form of transfinite induction (Proposition 4.13(a)). First, $\emptyset \in X$, since $G'\emptyset = \omega$. Second, assume $\alpha \in X$. We must show that $\alpha' \in X$. Since $\alpha \in X, G'\alpha$ is an infinite initial ordinal such that $(\forall\beta)(\beta <_0 \alpha \Rightarrow G'\beta <_0 G'\alpha)$. By definition, $G'(\alpha') = \mathscr{H}'(G'\alpha)$, the least initial ordinal $>_0 G'(\alpha)$. Assume $\beta <_o \alpha'$. Then $\beta <_0 \alpha \vee \beta = \alpha$. If $\beta <_0 \alpha$, then, since $\alpha \in X$, $G'\beta <_0 G'\alpha <_0 G'(\alpha')$. If $\beta = \alpha$, then $G'\beta = G'\alpha <_0 G'(\alpha')$. In either case, $G'\beta <_0 G'(\alpha')$, Hence, $\alpha' \in X$. Finally, assume $\mathrm{Lim}(\alpha) \wedge (\forall\beta)(\beta <_0 \alpha \Rightarrow \beta \in X)$. We must show that $\alpha \in X$. By definition, $G'\alpha = \bigcup(G''(\alpha))$. Now consider any $\beta <_0 \alpha$. Since $\mathrm{Lim}(\alpha), \beta' <_0 \alpha$. By assumption, $\beta' \in X$, that is, $G'(\beta')$ is an infinite initial ordinal such that, for any $\gamma <_0 \beta', G'\gamma <_0 G'(\beta')$. It follows that $G''(\alpha)$ is a non-empty set of ordinals without a maximum and, therefore, by Proposition 4.12, $G'\alpha$, which is $\bigcup(G''(\alpha))$, is a limit ordinal that is the least upper bound of $G''(\alpha)$. To conclude that $G'\alpha \in X$, we must show that $G'\alpha$ is an initial ordinal. For the sake of contradiction, assume that there exist δ such that $\delta <_0 G'(\alpha)$ and $\delta \cong G'\alpha$. Since $G'\alpha$ is the least upper bound of $G''(\alpha)$, there must exist some μ in $G''(\alpha)$ such that $\delta <_0 \mu$. Say, $\mu = G'\beta$ with $\beta <_0 \alpha$. So, $\delta \subseteq \mu = G'\beta \subset G'(\beta') \subseteq G'\alpha \cong \delta$. Since $\delta \subset G'(\beta'), \delta \in G'(\beta')$ and $\delta \preccurlyeq G'(\beta')$. On the other hand, since $G'(\beta') \subseteq G'\alpha \cong \delta, G'(\beta') \preccurlyeq \delta$. By Bernstein's theorem, $\delta \cong G'(\beta')$, contradicting the fact that $G'(\beta')$ is an initial ordinal.

(b) This follows from Corollary 4.16 and part (a).

(c) Assume, for the sake of contradiction, that there is an infinite initial ordinal that is not in the range of G, and let σ be the least such. By part (b), $\sigma \leqslant_o G`\sigma$ and, by part (a), $G`\sigma$ is an initial ordinal. Since σ is not in the range of G, $\sigma <_o G`\sigma$. Let $\mu$ be the least ordinal such that $\sigma <_o G`\mu$. Clearly, $\mu \neq \emptyset$, since $G`\emptyset = \omega <_o \sigma$. Assume first that $\mu$ is a successor ordinal $\gamma'$. Then, by the minimality of $\mu$, $G`\gamma <_o \sigma$. Since $G`(\gamma') = \mathscr{H}`(G`\gamma), G`(\gamma')$ is the least initial ordinal greater than $G`\gamma$. However, this contradicts the fact that $\sigma$ is an initial ordinal greater than $G`\gamma$ and $\sigma <_o G`(\gamma')$. So, $\mu$ must be a limit ordinal. Since $G`\mu = \bigcup(G``(\mu))$, the least upper bound of $G``(\mu)$, and $\sigma <_o G`\mu$, there is some $\delta <_o \mu$ such that $\sigma <_o G`\delta <_o G`\mu$, contradicting the minimality of μ.

Thus, by Proposition 4.33, G is a one–one $<_o$-preserving function from On onto the class of all infinite initial ordinals.

NOTATION

ω_α for $G`\alpha$

Hence, (a) $\omega_\emptyset = \omega$; (b) $\omega_{\alpha'}$ is the least initial ordinal greater than ω_α; (c) for a limit ordinal λ, ω_λ is the initial ordinal that is the least upper bound of the set of all ω_γ with $\gamma <_o \lambda$. Moreover, $\omega_\alpha \geq_o \alpha$ for all α. In addition, any infinite ordinal α is equinumerous with a unique initial ordinal $\omega_\beta \leqslant_o \alpha$, namely, with the least ordinal equinumerous with α.

Let us return now to ordinal arithmetic. We already have defined ordinal addition, multiplication and exponentiation (see Examples 1–2 on pages 249–50 and Exercise 4.36).

PROPOSITION 4.34

The following wfs are theorems.

(a) $\beta +_o 1 = \beta'$
(b) $\emptyset +_o \beta = \beta$
(c) $\emptyset <_o \beta \Rightarrow (\alpha <_o \alpha +_o \beta \wedge \beta \leqslant_o \alpha +_o \beta)$
(d) $\beta <_o \gamma \Rightarrow \alpha +_o \beta <_o \alpha +_o \gamma$
(e) $\alpha +_o \beta = \alpha +_o \delta \Rightarrow \beta = \delta$
(f) $\alpha <_o \beta \Rightarrow (\exists_1 \delta)(\alpha +_o \delta = \beta)$
(g) $\emptyset \neq x \subseteq On \Rightarrow \alpha +_o \bigcup_{\beta \in x} \beta = \bigcup_{\beta \in x} (\alpha +_o \beta)$
(h) $\emptyset <_o \alpha \wedge 1 <_o \beta \Rightarrow \alpha <_o \alpha \times_o \beta$
(i) $\emptyset <_o \alpha \wedge \emptyset <_o \beta \Rightarrow \alpha \leqslant_o \alpha \times_o \beta$
(j) $\gamma <_o \beta \wedge \emptyset <_o \alpha \Rightarrow \alpha \times_o \gamma <_o \alpha \times_o \beta$
(k) $x \subseteq On \Rightarrow \alpha \times_o \bigcup_{\beta \in x} \beta = \bigcup_{\beta \in x} (\alpha \times_o \beta)$

Proof

(a) $\beta +_o 1 = \beta +_o (\emptyset') = (\beta +_o \emptyset)' = \beta'$

(b) Prove $\emptyset +_o \beta = \beta$ by transfinite induction (Proposition 4.13(a)). Let $X = \{\beta \mid \emptyset +_o \beta = \beta\}$. First, $\emptyset \in X$, since $\emptyset +_o \emptyset = \emptyset$. If $\emptyset +_o \gamma = \gamma$, then $\emptyset +_o \gamma' = (\emptyset +_o \gamma)' = \gamma'$. If Lim($\alpha$) and $\emptyset +_o \tau = \tau$ for all $\tau <_o \alpha$, then $\emptyset +_o \alpha = \bigcup_{\tau <_o \alpha} (\emptyset +_o \tau) = \bigcup_{\tau <_o \alpha} \tau = \alpha$, since $\bigcup_{\tau <_o \alpha} \tau$ is the least upper bound of the set of all $\tau <_o \alpha$, which is α.

(c) Let $X = \{\beta \mid \emptyset <_o \beta \Rightarrow \alpha <_o \alpha +_o \beta\}$. Prove $X = On$ by transfinite induction. Clearly, $\emptyset \in X$. If $\gamma \in X$, then $\alpha \leqslant_o \alpha +_o \gamma$; hence $\alpha \leqslant_o \alpha +_o \gamma <_o (\alpha +_o \gamma)' = \alpha +_o \gamma'$. If Lim($\lambda$) and $\tau \in X$ for all $\tau <_o \lambda$, then $\alpha <_o \alpha' = \alpha +_o 1 \leqslant_o \bigcup_{\tau <_o \lambda} (\alpha +_o \tau) = \alpha +_o \lambda$. The second part is left as an exercise.

(d) Let $X = \{\gamma \mid (\forall\alpha)(\forall\beta)\ (\beta <_o \gamma \Rightarrow \alpha +_o \beta <_o \alpha +_o \gamma)\}$ and use transfinite induction. Clearly, $\emptyset \in X$. Assume $\gamma \in X$ and $\beta <_o \gamma'$. Then $\beta <_o \gamma$ or $\beta = \gamma$. If $\beta <_o \gamma$ then, since $\gamma \in X$, $\alpha +_o \beta <_o \alpha +_o \gamma <_o (\alpha +_o \gamma)' = \alpha +_o \gamma'$. If $\beta = \gamma$, then $\alpha +_o \beta = \alpha +_o \gamma <_o (\alpha +_o \gamma)' = \alpha +_o \gamma'$. Hence, $\gamma' \in X$. Assume Lim (λ) and $\tau \in X$ for all $\tau <_o \lambda$. Assume $\beta <_o \lambda$. Then $\beta <_o \tau$ for some $\tau <_o \lambda$, since Lim(λ). Hence, since $\tau \in X$, $\alpha +_o \beta <_o \alpha +_o \tau \leqslant_o \bigcup_{\tau <_o \lambda} (\alpha +_o \tau) = \alpha +_o \lambda$. Hence, $\lambda \in X$.

(e) Assume $\alpha +_o \beta = \alpha +_o \delta$. Now, either $\beta <_o \delta$ or $\delta <_o \beta$ or $\delta = \beta$. If $\beta <_o \delta$, then $\alpha +_o \beta <_o \alpha +_o \delta$ by part (d), and, if $\delta <_o \beta$, then $\alpha +_o \delta <_o \alpha +_o \beta$ by part (d); in either case, we get a contradiction with $\alpha +_o \beta = \alpha +_o \delta$. Hence, $\delta = \beta$.

(f) The uniqueness follows from part (e). Prove the existence by induction on β. Let $X = \{\beta \mid \alpha <_o \beta \Rightarrow (\exists_1\delta)(\alpha +_o \delta = \beta)\}$. Clearly, $\emptyset \in X$. Assume $\gamma \in X$ and $\alpha <_o \gamma'$. Hence, $\alpha = \gamma$ or $\alpha <_o \gamma$. If $\alpha = \gamma$, then $(\exists\delta)(\alpha +_o \delta = \gamma')$, namely, $\delta = 1$. If $\alpha <_o \gamma$, then, since $\gamma \in X$, $(\exists_1\delta)(\alpha +_o \delta = \gamma)$. Take an ordinal σ such that $\alpha +_o \sigma = \gamma$. Then $\alpha +_o \sigma' = (\alpha +_o \sigma)' = \gamma'$; thus, $(\exists\delta)(\alpha +_o \delta = \gamma')$; hence, $\gamma' \in X$. Assume now that Lim(λ) and $\tau \in X$ for all $\tau <_o \lambda$. Assume $\alpha <_o \lambda$. Now define a function f such that, for $\alpha <_o \mu <_o \lambda$, $f`\mu$ is the unique ordinal $\delta$ such that $\alpha +_o \delta = \mu$. But $\lambda = \bigcup_{\alpha <_o \mu <_o \lambda} \mu = \bigcup_{\alpha <_o \mu <_o \lambda} (\alpha +_o f`\mu)$. Let $\rho = \bigcup_{\alpha <_o \mu <_o \lambda} (f`\mu)$. Notice that, if $\alpha <_o \mu <_o \lambda$, then $f`\mu <_o f`(\mu')$; hence, $\rho$ is a limit ordinal. Then $\lambda = \bigcup_{\alpha <_o \mu <_o \lambda} (\alpha +_o f`\mu) = \bigcup_{\sigma <_o \rho} (\alpha +_o \sigma) = \alpha +_o \rho$.

(g) Assume $\emptyset \neq x \subseteq On$. By part (f), there is some δ such that $\alpha +_o \delta = \bigcup_{\beta \in x} (\alpha +_o \beta)$. We must show that $\delta = \bigcup_{\beta \in x} \beta$. If $\beta \in x$, then $\alpha +_o \beta \leqslant_o \alpha +_o \delta$. Hence, $\beta \leqslant_o \delta$ by part (d). Therefore, δ is an upper bound of the set of all β in x. So, $\bigcup_{\beta \in x} \beta \leqslant_o \delta$. On the other hand, if $\beta \in x$, then $\alpha +_o \beta \leqslant_o \alpha +_o \bigcup_{\beta \in x} \beta$. Hence, $\alpha +_o \delta = \bigcup_{\beta \in x} (\alpha +_o \beta) \leqslant_o \alpha +_o \bigcup_{\beta \in x} \beta$. Hence, $\alpha +_o \delta = \bigcup_{\beta \in x} (\alpha +_o \beta) \leqslant_o \alpha +_o \bigcup_{\beta \in x} \beta$ and so, by part (d), $\delta \leqslant_o \bigcup_{\beta \in x} \beta$. Therefore, $\delta = \bigcup_{\beta \in x} \beta$.

(h)–(k) are left as exercises.

PROPOSITION 4.35 The following wfs are theorems.

(a) $\beta \times_o 1 = \beta \wedge 1 \times_o \beta = \beta$
(b) $\emptyset \times_o \beta = \emptyset$
(c) $(\alpha +_o \beta) +_o \gamma = \alpha +_o (\beta +_o \gamma)$
(d) $(\alpha \times_o \beta) \times_o \gamma = \alpha \times_o (\beta \times_o \gamma)$
(e) $\alpha \times_o (\beta +_o \gamma) = (\alpha \times_o \beta) +_o (\alpha \times_o \gamma)$
(f) $\exp(\beta, 1) = \beta \wedge \exp(1, \beta) = 1$
(g) $\exp(\exp(\beta, \gamma), \delta) = \exp(\beta, \gamma \times_o \delta)$
(h) $\exp(\beta, \gamma +_o \delta) = \exp(\beta, \gamma) \times_o \exp(\beta, \delta)$[†]
(i) $\alpha >_o 1 \wedge \beta <_o \gamma \Rightarrow \exp(\alpha, \beta) <_o \exp(\alpha, \gamma)$

Proof

(a) $\beta \times_o 1 = \beta \times_o \emptyset' = (\beta \times_o \emptyset) +_o \beta = \emptyset +_o \beta = \beta$, by Proposition 4.34(b). Prove $1 \times_o \beta = \beta$ by transfinite induction.
(b) Prove $\emptyset \times_o \beta = \emptyset$ by transfinite induction.
(c) Let $X = \{\gamma \mid (\forall\alpha)(\forall\beta)((\alpha +_o \beta) +_o \gamma = \alpha +_o (\beta +_o \gamma))\}$. $\emptyset \in X$, since $(\alpha +_o \beta) +_o \emptyset = (\alpha +_o \beta) = \alpha +_o (\beta +_o \emptyset)$. Now assume $\gamma \in X$. Then $(\alpha +_o \beta) +_o \gamma' = ((\alpha +_o \beta) +_o \gamma)' = (\alpha +_o (\beta +_o \gamma))' = \alpha +_o (\beta +_o \gamma)' = \alpha +_o (\beta +_o \gamma')$. Hence, $\gamma' \in X$. Assume now that $\mathrm{Lim}(\lambda)$ and $\tau \in X$ for all $\tau <_o \lambda$. Then $(\alpha +_o \beta) +_o \lambda = \bigcup_{\tau <_o \lambda}((\alpha +_o \beta) +_o \tau) = \bigcup_{\tau <_o \lambda}(\alpha +_o (\beta +_o \tau)) = \alpha +_o \bigcup_{\tau <_o \lambda}(\beta +_o \tau)$, by Proposition 4.34(g), and this is equal to $\alpha +_o (\beta +_o \lambda)$.

(d)–(i) are left as exercises.

We would like to consider for a moment the properties of ordinal addition, multiplication and exponentiation when restricted to ω.

PROPOSITION 4.36
Assume α, β, γ are in ω. Then:

(a) $\alpha +_o \beta \in \omega$
(b) $\alpha \times_o \beta \in \omega$
(c) $\exp(\alpha, \beta) \in \omega$
(d) $\alpha +_o \beta = \beta +_o \alpha$
(e) $\alpha \times_o \beta = \beta \times_o \alpha$
(f) $(\alpha +_o \beta) \times_o \gamma = (\alpha \times_o \gamma) +_o (\beta \times_o \gamma)$
(g) $\exp(\alpha \times_o \beta, \gamma) = \exp(\alpha, \gamma) \times_o \exp(\beta, \gamma)$

[†]In traditional notation, the results of (f)–(h) would be written as $\beta^1 = \beta$, $1^\beta = 1$, $(\beta^\gamma)^\delta = \beta^{\gamma \times_o \delta}$, $\beta^{\gamma +_o \delta} = \beta^\gamma \times_o \beta^\delta$.

Proof

(a) Use induction up to ω (Proposition 4.13(c)). Let $X = \{\beta \mid \beta \in \omega \wedge (\forall\alpha)(\alpha \in \omega) \Rightarrow \alpha +_o \beta \in \omega)\}$. Clearly, $\emptyset \in X$. Assume $\beta \in X$. Consider any $\alpha \in \omega$. Then $\alpha +_o \beta \in \omega$. Hence, $\alpha +_o \beta' = (\alpha +_o \beta)' \in \omega$ by Proposition 4.11(a). Thus, $\beta' \in X$.

(b) and (c) are left as exercises.

(d) *Lemma.* $\vdash \alpha \in \omega \wedge \beta \in \omega \Rightarrow \alpha' +_o \beta = \alpha +_o \beta'$. Let $Y = \{\beta \mid \beta \in \omega \wedge (\forall\alpha)\ (\alpha \in \omega \Rightarrow \alpha' +_o \beta = \alpha +_o \beta')\}$. Clearly, $\emptyset \in Y$. Assume $\beta \in Y$. Consider any $\alpha \in \omega$. So, $\alpha' +_o \beta = \alpha +_o \beta'$. Then $\alpha' +_o \beta' = (\alpha' +_o \beta)' = (\alpha +_o \beta')' = \alpha +_o (\beta')'$. Hence, $\beta' \in Y$.

To prove (d), let $X = \{\beta \mid \beta \in \omega \wedge (\forall\alpha)(\alpha \in \omega \Rightarrow \alpha +_o \beta = \beta +_o \alpha)\}$. Then $\emptyset \in X$ and it is easy to prove, using the lemma, that $\beta \in X \Rightarrow \beta' \in X$.

(e)–(g) are left as exercises.

The reader will have noticed that we have not asserted for ordinals certain well-known laws, such as the commutative laws for addition and multiplication, that hold for other familiar number systems. In fact, these laws fail for ordinals, as the following examples show.

Examples

1. $(\exists\alpha)(\exists\beta)(\alpha +_o \beta \neq \beta +_o \alpha)$

$$1 +_o \omega = \bigcup_{\alpha <_o \omega} (1 +_o \alpha) = \omega$$

$$\omega +_o 1 = \omega' >_o \omega$$

2. $(\exists\alpha)(\exists\beta)(\alpha \times_o \beta \neq \beta \times_o \alpha)$

$$2 \times_o \omega = \bigcup_{\alpha <_o \omega} (2 \times_o \alpha) = \omega$$

$$\omega \times_o 2 = \omega \times_o (1 +_o 1) = (\omega \times_o 1) +_o (\omega \times_o 1) = \omega +_o \omega >_o \omega$$

3. $(\exists\alpha)(\exists\beta)(\exists\gamma)((\alpha +_o \beta) \times_o \gamma \neq (\alpha \times_o \gamma) +_o (\beta \times_o \gamma))$

$$(1 +_o 1) \times_o \omega = 2 \times_o \omega = \omega$$

$$(1 \times_o \omega) +_o (1 \times_o \omega) = \omega +_o \omega >_o \omega$$

4. $(\exists\alpha)(\exists\beta)(\exists\gamma)(\exp(\alpha \times_o \beta, \gamma) \neq \exp(\alpha, \gamma) \times_o \exp(\beta, \gamma))$

$$\exp(2 \times_o 2, \omega) = \exp(4, \omega) = \bigcup_{\alpha <_o \omega} \exp(4, \alpha) = \omega$$

$$\exp(2, \omega) = \bigcup_{\alpha <_o \omega} \exp(2, \alpha) = \omega$$

So, $\exp(2, \omega) \times_o \exp(2, \omega) = \omega \times_o \omega >_o \omega$.

Given any wf $\mathscr{B}$ of formal number theory S (see Chapter 3), we can associate with $\mathscr{B}$ a wf $\mathscr{B}^*$ of NBG as follows: first, replace every '$+$' by '$+_o$', every '$\cdot$' by '$\times_o$', and every '$f_1^1(t)$' by '$t \cup \{t\}$'[†]; then, if $\mathscr{B}$ is $\mathscr{C} \Rightarrow \mathscr{D}$ or $\neg\mathscr{C}$,

[†]In abbreviated notation for S, '$f_1^1(t)$' is written as t', and in abbreviated notation in NBG, '$t \cup \{t\}$' is written as t'. So, no change will take place in these abbreviated notations.

respectively, and we have already found $\mathscr{C}^*$ and $\mathscr{D}^*$, let $\mathscr{B}^*$ be $\mathscr{C}^* \Rightarrow \mathscr{D}^*$ or $\neg\mathscr{C}^*$, respectively; if $\mathscr{B}$ is $(\forall x)\mathscr{C}(x)$, replace it by $(\forall x)(x \in \omega \Rightarrow \mathscr{C}^*(x))$. This completes the definition of $\mathscr{B}^*$. Now, if $x_1, \ldots, x_n$ are the free variables (if any) of $\mathscr{B}$, prefix $(x_1 \in \omega \wedge \ldots \wedge x_n \in \omega) \Rightarrow$ to $\mathscr{B}^*$, obtaining a wf $\mathscr{B}\#$. This amounts to restricting all variables to ω and interpreting addition, multiplication and the successor function on natural numbers as the corresponding operations on ordinals. Then every axiom $\mathscr{B}$ of S is transformed into a theorem $\mathscr{B}\#$ of NBG. (Axioms (S1)–(S3) are obviously transformed into theorems, (S4)# is a theorem by Proposition 4.10(c), and (S5)#–(S8)# are properties of ordinal addition and multiplication.) Now, for any wf $\mathscr{B}$ of S, $\mathscr{B}\#$ is predicative. Hence, by Proposition 4.4, all instances of (S9)# are provable by Proposition 4.13(c). (In fact, assume $\mathscr{B}\#(\emptyset) \wedge (\forall x)(x \in \omega \Rightarrow (\mathscr{B}\#(x) \Rightarrow \mathscr{B}\#(x')))$. Let $X = \{y \mid y \in \omega \wedge \mathscr{B}\#(y)\}$. Then, by Proposition 4.13(c), $(\forall x)(x \in \omega \Rightarrow \mathscr{B}\#(x))$.) Applications of modus ponens are easily seen to be preserved under the transformation of $\mathscr{B}$ into $\mathscr{B}\#$. As for the generalization rule, consider a wf $\mathscr{B}(x)$ and assume that $\mathscr{B}\#(x)$ is provable in NBG. But $\mathscr{B}\#(x)$ is of the form $x \in \omega \wedge y_1 \in \omega \wedge \ldots \wedge y_n \in \omega \Rightarrow \mathscr{B}^*(x)$. Hence, $y_1 \in \omega \wedge \ldots \wedge y_n \in \omega \Rightarrow (\forall x)(x \in \omega \Rightarrow \mathscr{B}^*(x))$ is provable in NBG. But this wf is just $((\forall x)\mathscr{B}(x))\#$. Hence, application of Gen leads from theorems to theorems. Therefore, for every theorem $\mathscr{B}$ of S, $\mathscr{B}\#$ is a theorem of NBG, and we can translate into NBG all the theorems of S proved in Chapter 3.

One can check that the number-theoretic function h such that, if x is the Gödel number of a wf $\mathscr{B}$ of S, then $h(x)$ is the Gödel number of $\mathscr{B}\#$, and if x is not the Gödel number of a wf of S, then $h(x) = 0$, is recursive (in fact, primitive recursive). Let K be any consistent extension of NBG. As we saw above, if x is the Gödel number of a theorem of S, then $h(x)$ is the Gödel number of a theorem of NBG and, hence, also a theorem of K. Let S(K) be the extension of S obtained by taking as axioms all wfs $\mathscr{B}$ of the language of S such that $\mathscr{B}\#$ is a theorem of K. Since K is consistent, S(K) must be consistent. Therefore, since S is essentially recursively undecidable (by Corollary 3.46), S(K) is recursively undecidable. Now, assume K is recursively decidable; that is, the set T_K of Gödel numbers of theorems of K is recursive. But $C_{T_{S(K)}}(x) = C_{T_K}(h(x))$ for any x, where $C_{T_{S(K)}}$ and C_{T_K} are the characteristic functions of $T_{S(K)}$ and T_K. Hence, $T_{S(K)}$ would be recursive, contradicting the recursive undecidability of S(K). Therefore, K is recursively undecidable, and thus, if NBG is consistent, NBG is essentially recursively undecidable. Recursive undecidability of a recursively axiomatizable theory implies incompleteness (see Proposition 3.47). Hence, NBG is also essentially incomplete. Thus, we have the following result: *if NBG is consistent, then NBG is essentially recursively undecidable and essentially incomplete.* (It is possible to prove this result directly in the same way that the corresponding result was proved for S in Chapter 3.)

Exercise

4.68 Prove that a predicate calculus with a single binary predicate letter is recursively undecidable. [*Hint*: Use Proposition 3.49 and the fact that NBG has a finite number of proper axioms.]

There are a few facts about the 'cardinal arithmetic' of ordinal numbers that we would like to deal with now. By 'cardinal arithmetic' we mean properties connected with the operations of union ($\bigcup$), Cartesian product (x) and X^Y, as opposed to the properties of $+_o$, $\times_o$ and exp. Observe that $\times$ is distinct from $\times_o$; also notice that ordinal exponentiation $\exp(\alpha, \beta)$ has nothing to do with X^Y, the class of all functions from Y into X. (From Example 4 on page 269 we see that $\exp(2, \omega)$ is ω, whereas, from Cantor's theorem, $\omega \prec 2^\omega$, where 2^ω is the set of functions from ω into 2.

PROPOSITION 4.37

(a) $\vdash \omega \times \omega \cong \omega$
(b) $\vdash 2 \preccurlyeq X \wedge 2 \preccurlyeq Y \Rightarrow X \cup Y \preccurlyeq X \times Y$
(c) $\vdash \mathrm{Den}(x) \wedge \mathrm{Den}(y) \Rightarrow \mathrm{Den}(x \cup y)$

Proof

(a) Let f be a function with domain ω such that, if $\alpha \in \omega$, then $f`\alpha = \langle \alpha, \emptyset \rangle$. Then f is a one–one function from ω into a subset of $\omega \times \omega$. Hence, $\omega \preccurlyeq \omega \times \omega$. Conversely, let g be a function with domain $\omega \times \omega$ such that, for any $\langle \alpha, \beta \rangle$ in $\omega \times \omega$, $g`\langle \alpha, \beta \rangle = \exp(2, \alpha) \times_o \exp(3, \beta)$. We leave it as an exercise to show that g is a one–one function from $\omega \times \omega$ into ω. Hence, $\omega \times \omega \preccurlyeq \omega$. So, by Bernstein's theorem, $\omega \times \omega \cong \omega$.

(b) Assume $a_1 \in X, a_2 \in X, a_1 \neq a_2, b_1 \in Y, b_2 \in Y, b_1 \neq b_2$. Define

$$f`x = \begin{cases} \langle a_1, b_1 \rangle & \text{if } x \in X \\ \langle a_1, x \rangle & \text{if } x \in Y - X \text{ and } x \neq b_1 \\ \langle a_2, b_2 \rangle & \text{if } x = b_1 \text{ and } x \in Y - X \end{cases}$$

Then f is a one-one function with domain $X \cup Y$ and range a subset of $X \times Y$. Hence, $X \cup Y \preccurlyeq X \times Y$.

(c) Assume $\mathrm{Den}(x)$ and $\mathrm{Den}(y)$. Hence, each of x and y contains at least two elements. Then, by part (b), $x \cup y \preccurlyeq x \times y$. But $x \cong \omega$ and $y \cong \omega$. Hence, $x \times y \cong \omega \times \omega$. Therefore, $x \cup y \preccurlyeq \omega \times \omega \cong \omega$. By Proposition 4.30, either Den $(x \cup y)$ or Fin $(x \cup y)$. But $x \subseteq x \cup y$ and $\mathrm{Den}(x)$; hence, $\neg\mathrm{Fin}(x \cup y)$.

For the further study of ordinal addition and multiplication, it is quite useful to obtain concrete interpretations of these operations.

PROPOSITION 4.38 (ADDITION)

Assume that $\langle r, x\rangle$ is similar to $\langle \mathrm{E}_\alpha, \alpha\rangle$, that $\langle s, y\rangle$ is similar to $\langle \mathrm{E}_\beta, \beta\rangle$, and that $x \cap y = \emptyset$. Let t be the relation on $x \cup y$ consisting of all $\langle u, v\rangle$ such that $\langle u, v\rangle \in x \times y$ or $u \in x \wedge v \in x \wedge \langle u, v\rangle \in r$ or $u \in y \wedge v \in y \wedge \langle u, v\rangle \in s$ (that is, t is the same as r in the set x, the same as s in the set y, and every element of x t-precedes every element of y). Then t is a well-ordering of $x \cup y$, and $\langle t, x \cup y\rangle$ is similar to $\langle \mathrm{E}_{\alpha +_o \beta}, \alpha +_o \beta\rangle$.

Proof

First, it is simple to verify that t is a well-ordering of $x \cup y$, since r is a well-ordering of x and s is a well-ordering of y. To show that $\langle t, x \cup y\rangle$ is similar to $\langle \mathrm{E}_{\alpha +_o \beta}, \alpha +_o \beta\rangle$, use transfinite induction on β. For $\beta = \emptyset$, $y = \emptyset$. Hence, $t = r, x \cup y = x$, and $\alpha +_o \beta = \alpha$. So, $\langle t, \alpha \cup \beta\rangle$ is similar to $\langle \mathrm{E}_{\alpha +_o \beta}, \alpha +_o \beta\rangle$. Assume the proposition for γ and let $\beta = \gamma'$. Since $\langle s, y\rangle$ is similar to $\langle \mathrm{E}_\beta, \beta\rangle$, we have a function f with domain y and range β such that, for any u, v in y, $\langle u, v\rangle \in s$ if and only if $f`u \in f`v$. Let $b = (\breve{f})`\gamma$, let $y_1 = y - \{b\}$ and let $s_1 = s \cap (y_1 \times y_1)$. Since b is the s-maximum of y, it follows easily that s_1 well-orders y_1. Also, $y_1 \lceil f$ is a similarity mapping of y_1 onto γ. Let $t_1 = t \cap ((x \cup y_1) \times (x \cup y_1))$. By inductive hypothesis, $\langle t_1, x \cup y_1\rangle$ is similar to $\langle \mathrm{E}_{\alpha +_o \gamma}, \alpha +_o \gamma\rangle$, by means of some similarity mapping g with domain $x \cup y_1$ and range $\alpha +_o \gamma$. Extend g to $g_1 = g \cup \{\langle b, \alpha +_o \gamma\rangle\}$, which is a similarity mapping of $x \cup y$ onto $(\alpha +_o \gamma)' = \alpha +_o \gamma' = \alpha +_o \beta$. Finally, if $\mathrm{Lim}(\beta)$ and our proposition holds for all $\tau <_o \beta$, assume that f is a similarity mapping of y onto β. Now, for each $\tau <_o \beta$, let $y_\tau = (\breve{f})``\tau, s_\tau = s \cap (y_\tau \times y_\tau)$, and $t_\tau = t \cap ((x \cup y_\tau) \times (x \cup y_\tau))$. By inductive hypothesis and Corollary 4.18(b), there is a unique similarity mapping g_τ of $\langle t_\tau, x \cup y_\tau\rangle$ with $\langle E_{\alpha +_o \tau}, \alpha +_o \tau\rangle$; also, if $\tau_1 <_o \tau_2 <_o \beta$, then, since $(x \cup y_{\tau_1}) \lceil g_{\tau_2}$ is a similarity mapping of $\langle t_{\tau_1}, x \cup y_{\tau_1}\rangle$ with $\langle \mathrm{E}_{\alpha +_o \tau_1}, \alpha +_o \tau_1\rangle$ and, by the uniqueness of g_{τ_1}, $(x \cup y_{\tau_1}) \lceil g_{\tau_2} = g_{\tau_1}$; that is, g_{τ_2} is an extension of g_{τ_1}. Hence, if $g = \bigcup_{\tau <_o \beta} g_\tau$ and $\lambda = \bigcup_{\tau <_o \beta} (\alpha +_o \tau)$, then g is a similarity mapping of $\langle t, \bigcup_{\tau <_o \beta} (x \cup y_\tau)\rangle$ with $\langle E_\lambda, \lambda\rangle$. But, $\bigcup_{\tau <_o \beta} (x \cup y_\tau) = x \cup y$ and $\bigcup_{\tau <_o \beta} (\alpha +_o \tau) = \alpha +_o \beta$. This completes the transfinite induction.

PROPOSITION 4.39 (MULTIPLICATION)

Assume that $\langle r, x\rangle$ is similar to $\langle \mathrm{E}_\alpha, \alpha\rangle$ and that $\langle s, y\rangle$ is similar to $\langle \mathrm{E}_\beta, \beta\rangle$. Let the relation t on $x \times y$ consist of all pairs $\langle\langle u, v\rangle, \langle w, z\rangle\rangle$ such that u and w are in x and v and z are in y, and either $\langle v, z\rangle \in s$ or $(v = z \wedge \langle u, w\rangle \in r)$. Then t is a well-ordering of $x \times y$ and $\langle t, x \times y\rangle$ is similar to $\langle \mathrm{E}_{\alpha \times_o \beta}, \alpha \times_o \beta\rangle$.[†]

[†]The ordering t is called an *inverse lexicographical ordering* because it orders pairs as follows: first, according to the size of their second components and then, if their second components are equal, according to the size of their first components.

Proof

This is left as an exercise. Proceed as in the proof of Proposition 4.38.

Examples

1. $2 \times_o \omega = \omega$. Let $\langle r, x\rangle = \langle E_2, 2\rangle$ and $\langle s, y\rangle = \langle E_\omega, \omega\rangle$. Then the Cartesian product $2 \times \omega$ is well-ordered as follows: $\langle\emptyset,\emptyset\rangle, \langle 1,\emptyset\rangle, \langle\emptyset, 1\rangle, \langle 1, 1\rangle, \langle\emptyset, 2\rangle, \langle 1, 2\rangle, \ldots, \langle\emptyset, n\rangle, \langle 1, n\rangle, \langle\emptyset, n+1\rangle, \langle 1, n+1\rangle, \ldots$
2. By Proposition 4.34(a), $2 = 1' = 1 +_o 1$. Then by Proposition 4.35(e,a), $\omega \times_o 2 = (\omega \times_o 1) +_o (\omega \times_o 1) = \omega +_o \omega$. Let $\langle r, x\rangle = \langle E_\omega, \omega\rangle$ and $\langle s, y\rangle = \langle E_2, 2\rangle$. Then the Cartesian product $\omega \times 2$ is well-ordered as follows: $\langle\emptyset,\emptyset\rangle, \langle 1,\emptyset\rangle, \langle 2,\emptyset\rangle, \ldots, \langle\emptyset, 1\rangle, \langle 1, 1\rangle, \langle 2, 1\rangle, \ldots$

PROPOSITION 4.40

For all α, $\omega_\alpha \times \omega_\alpha \cong \omega_\alpha$.

Proof

(Sierpinski, 1958) Assume this is false and let α be the least ordinal such that $\omega_\alpha \times \omega_\alpha \cong \omega_\alpha$ is false. Then $\omega_\beta \times \omega_\beta \cong \omega_\beta$ for all $\beta <_o \alpha$. By Proposition 4.37(a), $\alpha >_o \emptyset$. Now let $P = \omega_\alpha \times \omega_\alpha$ and, for $\beta <_o \omega_\alpha$, let $P_\beta = \{\langle\gamma, \delta\rangle | \gamma +_o \delta = \beta\}$. First we wish to show that $P = \bigcup_{\beta <_o \omega_\alpha} P_\beta$. Now, if $\gamma +_o \delta = \beta <_o \omega_\alpha$, then $\gamma \leqslant_o \beta <_o \omega_\alpha$ and $\delta \leqslant_o \beta <_o \omega_\alpha$; hence, $\langle\gamma, \delta\rangle \in \omega_\alpha \times \omega_\alpha = P$. Thus, $\bigcup_{\beta <_o \omega_\alpha} P_\beta \subseteq P$. To show that $P \subseteq \bigcup_{\beta <_o \omega_\alpha} P_\beta$, it suffices to show that, if $\gamma <_o \omega_\alpha$ and $\delta <_o \omega_\alpha$, then $\gamma +_o \delta <_o \omega_\alpha$. This is clear when γ or δ is finite. Hence, we may assume that γ and δ are equinumerous with initial ordinals $\omega_\sigma \leqslant_o \gamma$ and $\omega_\rho \leqslant_o \delta$, respectively. Let ζ be the larger of σ and ρ. Since $\gamma <_o \omega_\alpha$ and $\delta <_o \omega_\alpha$, then $\omega_\zeta <_o \omega_\alpha$. Hence, by the minimality of $\alpha, \omega_\zeta \times \omega_\zeta \cong \omega_\zeta$. Let $x = \gamma \times \{\emptyset\}$ and $y = \delta \times \{1\}$. Then, by Proposition 4.38, $x \cup y \cong \gamma +_o \delta$. Since $\gamma \cong \omega_\sigma$ and $\delta \cong \omega_\rho, x \cong \omega_\sigma \times \{\emptyset\}$ and $y \cong \omega_\rho \times \{1\}$. Hence, since $x \cap y = \emptyset, x \cup y \cong (\omega_\sigma \times \{\emptyset\}) \cup (\omega_\rho \times \{1\})$. But, by Proposition 4.37(b), $(\omega_\sigma \times \{\emptyset\}) \cup (\omega_\rho \times \{1\}) \preccurlyeq (\omega_\sigma \times \{\emptyset\}) \times (\omega_\rho \times 1\}) \cong \omega_\sigma \times \omega_\rho \preccurlyeq \omega_\zeta \times \omega_\zeta \cong \omega_\zeta$. Hence, $\gamma +_o \delta \preccurlyeq \omega_\zeta <_o \omega_\alpha$. It follows that $\gamma +_o \delta <_o \omega_\alpha$. (If $\omega_\alpha \leqslant_o \gamma +_o \delta$, then $\omega_\alpha \preccurlyeq \omega_\zeta$. Since $\omega_\zeta <_o \omega_\alpha$, $\omega_\zeta \preccurlyeq \omega_\alpha$. So, by Bernstein's theorem, $\omega_\alpha \cong \omega_\zeta$, contradicting the fact that ω_α is an initial ordinal.) Thus, $P = \bigcup_{\beta <_o \omega_\alpha} P_\beta$. Consider P_β for any $\beta <_o \omega_\alpha$. By Proposition 4.34(f), for each $\gamma \leqslant_o \beta$, there is exactly one ordinal δ such that $\gamma +_o \delta = \beta$. Hence, there is a similarity mapping from β' onto P_β, where P_β is ordered according to the size of the first component γ of the pairs $\langle\gamma, \delta\rangle$. Define the following relation R on P. For any $\gamma <_o \omega_\alpha, \delta <_o \omega_\alpha$, $\mu <_o \omega_\alpha, \nu <_o \omega_\alpha$, $\langle\langle\gamma, \delta\rangle, \langle\mu, \nu\rangle\rangle \in R$ if and only if either $\gamma +_o \delta <_o \mu +_o \nu$ or $(\gamma +_o \delta = \mu +_o \nu \wedge \gamma <_o \mu)$. Thus, if $\beta_1 <_o \beta_2 <_0 \omega_\alpha$, then the pairs in P_{β_1} R-precede the

pairs in P_{β_2}, and, within each P_β, the pairs are R-ordered according to the size of their first components. One easily verifies that R well-orders P. Since $P = \omega_\alpha \times \omega_\alpha$, it suffices now to show that $\langle R, P\rangle$ is similar to $\langle E_{\omega_\alpha}, \omega_\alpha\rangle$. By Proposition 4.19, $\langle R, P\rangle$ is similar to some $\langle E_\xi, \xi\rangle$, where ξ is an ordinal. Hence, $P \cong \xi$. Assume that $\xi >_o \omega_\alpha$. There is a similarity mapping f between $\langle E_\xi, \xi\rangle$ and $\langle R, P\rangle$. Let $b = f`(\omega_\alpha)$; then b is an ordered pair $\langle \gamma, \delta\rangle$ with $\gamma <_o \omega_\alpha, \delta <_o \omega_\alpha$, and $\omega_\alpha \downarrow f$ is a similarity mapping between $\langle E_{\omega_\alpha}, \omega_\alpha\rangle$ and the R-segment $Y = \mathrm{Seg}_R(P, \langle \gamma, \delta\rangle)$ of P determined by $\langle \gamma, \delta\rangle$. Then $Y \cong \omega_\alpha$. If we let $\beta = \gamma +_o \delta$, then, if $\langle \sigma, \rho\rangle \in Y$, we have $\sigma +_o \rho \leqslant_o \gamma +_o \delta = \beta$; hence, $\sigma \leqslant_o \beta$ and $\rho \leqslant_o \beta$. Therefore, $Y \subseteq \beta' \times \beta'$. But $\beta' <_o \omega_\alpha$. Since β is obviously not finite, $\beta' \cong \omega_\mu$ with $\mu <_o \alpha$. By the minimality of α, $\omega_\mu \times \omega_\mu \cong \omega_\mu$. So, $\omega_\alpha \cong Y \preccurlyeq \omega_\mu$, contradicting $\omega_\mu \prec \omega_\alpha$. Thus, $\xi \leqslant_o \omega_\alpha$ and, therefore, $P \preccurlyeq \omega_\alpha$. Let h be the function with domain ω_α such that $h`\beta = \langle \beta, \emptyset\rangle$ for every $\beta <_o \omega_\alpha$. Then h is one–one correspondence between ω_α and the subset $\omega_\alpha \times \{\emptyset\}$ of P and, therefore, $\omega_\alpha \preccurlyeq P$. By Bernstein's theorem, $\omega_\alpha \cong P$, contradicting the definition of α. Hence, $\omega_\beta \times \omega_\beta \cong \omega_\beta$ for all β.

COROLLARY 4.41

If $x \cong \omega_\alpha$ and $y \cong \omega_\beta$, and if γ is the maximum of α and β, then $x \times y \cong \omega_\gamma$ and $x \cup y \cong \omega_\gamma$. In particular, $\omega_\alpha \times \omega_\beta \cong \omega_\gamma$.

Proof

By Propositions 4.40 and 4.37(b), $\omega_\gamma \preccurlyeq x \cup y \preccurlyeq x \times y \cong \omega_\alpha \times \omega_\beta \preccurlyeq \omega_\gamma \times \omega_\gamma \cong \omega_\gamma$. Hence, by Bernstein's theorem, $x \times y \cong \omega_\gamma$ and $x \cup y \cong \omega_\gamma$.

Exercises

4.69 Prove that the following are theorems of NBG.

(a) $x \preccurlyeq \omega_\alpha \Rightarrow x \cup \omega_\alpha \cong \omega_\alpha$

(b) $\omega_\alpha +_c \omega_\alpha \cong \omega_\alpha$

(c) $\emptyset \neq x \preccurlyeq \omega_\alpha \Rightarrow x \times \omega_\alpha \cong \omega_\alpha$

(d) $\emptyset \neq x \prec \omega \Rightarrow (\omega_\alpha)^x \cong \omega_\alpha$

4.70. Prove that the following are theorems of NBG

(a) $\mathscr{P}(\omega_\alpha) \times \mathscr{P}(\omega_\alpha) \cong \mathscr{P}(\omega_\alpha)$

(b) $x \preccurlyeq \mathscr{P}(\omega_\alpha) \Rightarrow x \cup \mathscr{P}(\omega_\alpha) \cong \mathscr{P}(\omega_\alpha)$

(c) $\emptyset \neq x \preccurlyeq \mathscr{P}(\omega_\alpha) \Rightarrow x \times \mathscr{P}(\omega_\alpha) \cong \mathscr{P}(\omega_\alpha)$

(d) $\emptyset \neq x \preccurlyeq \omega_\alpha \Rightarrow (\mathscr{P}(\omega_\alpha))^x \cong \mathscr{P}(\omega_\alpha)$

(e) $1 \prec x \preccurlyeq \omega_\alpha \Rightarrow x^{\omega_\alpha} \cong (\omega_\alpha)^{\omega_\alpha} \cong (\mathscr{P}(\omega_\alpha))^{\omega_\alpha} \cong \mathscr{P}(\omega_\alpha)$

4.71 Assume $y \neq \emptyset \wedge y \cong y +_c y$. (This assumption holds for $y = \omega_\alpha$ by Corollary 4.41 and for $y = \mathscr{P}(\omega_\alpha)$ by Exercise 4.70(b). It will turn out to hold

for all infinite sets y if the axiom of choice holds.) Prove the following properties of y.

(a) $\mathrm{Inf}(y)$

(b) $y \cong 1 +_c y$

(c) $(\exists u)(\exists v)(y = u \cup v \wedge u \cap v = \emptyset \wedge u \cong y \wedge v \cong y)$

(d) $\{z | z \subseteq y \wedge z \cong y\} \cong \mathscr{P}(y)$

(e) $\{z | z \subseteq y \wedge \mathrm{Inf}(z)\} \cong \mathscr{P}(y)$

(f) $(\exists f)(y \underset{f}{\cong} y \wedge (\forall u)(u \in y \Rightarrow f`u \neq u))$

4.72 Assume $y \cong y \times y \wedge 1 \prec y$. (This holds when $y = \omega_\alpha$ by Proposition 4.40 and for $y = \mathscr{P}(\omega_\alpha)$ by Exercise 4.70(a). It is true for all infinite sets y if the axiom of choice holds.) Prove the following properties of y.

(a) $y \cong y +_c y$

(b)[D] Let $\mathrm{Perm}(y)$ denote $\{f | y \underset{f}{\cong} y\}$. Then $\mathrm{Perm}(y) \cong \mathscr{P}(y)$.

4.5 THE AXIOM OF CHOICE. THE AXIOM OF REGULARITY

The axiom of choice is one of the most celebrated and contested statements of the theory of sets. We shall state it in the next proposition and show its equivalence to several other important assertions.

PROPOSITION 4.42

The following wfs are equivalent.

(a) *Axiom of choice* (AC). For any set x, there is a function f such that, for any non-empty subset y of x, $f`y \in y$. (f is called a *choice function* for x.)

(b) *Multiplicative axiom* (Mult). If x is a set of pairwise disjoint non-empty sets, then there is a set y (called a *choice set* for x) such that y contains exactly one element of each set in x:

$$(\forall u)(u \in x \Rightarrow u \neq \emptyset \wedge (\forall v)(v \in x \wedge v \neq u \Rightarrow v \cap u = \emptyset)) \Rightarrow$$
$$(\exists y)(\forall u)(u \in x \Rightarrow (\exists_1 w)(w \in u \cap y))$$

(c) *Well-ordering principle* (WO). Every set can be well-ordered: $(\forall x)(\exists y)(y \text{ We } x)$.

(d) *Trichotomy* (Trich). $(\forall x)(\forall y)(x \preccurlyeq y \vee y \preccurlyeq x)$[†]

[†]This is equivalent to $(\forall x)(\forall y)(x \prec y \vee x \cong y \vee y \prec x)$, which explains the name 'trichotomy' for this principle.

(e) *Zorn's Lemma* (Zorn). Any non-empty partially ordered set x, in which every chain (i.e., every totally ordered subset) has an upper bound, has a maximal element:

$$(\forall x)(\forall y)([(y \text{ Part } x) \wedge (\forall u)(u \subseteq x \wedge y \text{ Tot } u \Rightarrow$$
$$(\exists v)(v \in x \wedge (\forall w)(w \in u \Rightarrow w = v \vee \langle w, v\rangle \in y)))] \Rightarrow$$
$$(\exists v)(v \in x \wedge (\forall w)(w \in x \Rightarrow \langle v, w\rangle \notin y)))$$

Proof

1. $\vdash$ WO $\Rightarrow$ Trich. Given sets x and y, then, by WO, x and y can be well-ordered. Hence, by Proposition 4.19, $x \cong \alpha$ and $y \cong \beta$ for some ordinals α and β. But, by Exercise 4.52, $\alpha \preccurlyeq \beta$ or $\beta \preccurlyeq \alpha$. Therefore, $x \preccurlyeq y$ or $y \preccurlyeq x$.

2. $\vdash$ Trich $\Rightarrow$ WO. Given a set x, Hartogs' theorem yields an ordinal α such that α is not equinumerous with any subset of x, that is, $\alpha \preccurlyeq x$ is false. So, by Trich, $x \preccurlyeq \alpha$, that is, x is equinumerous with some subset y of α. Hence, by translating the well-ordering E_y of y to x, x can be well-ordered.

3. $\vdash$ WO$\Rightarrow$ Mult. Let x be a set of non-empty pairwise disjoint sets. By WO, there is a well-ordering R of $\bigcup x$. Hence, there is a function f with domain x such that, for any u in x, f'u is the R-least element of u. (Notice that u is a subset of $\bigcup x$.)

4. $\vdash$ Mult $\Rightarrow$ AC. For any set x, we can define a one–one function g such that, for each non-empty subset u of x, g'$u = u \times \{u\}$. Let x_1 be the range of g. Then x_1 is a set of non-empty pairwise disjoint sets. Hence, by Mult, there is a choice set y for x_1. Therefore, if u is a non-empty subset of x, then $u \times \{u\}$ is in x_1, and so y contains exactly one element $\langle v, u\rangle$ in $u \times \{u\}$. Then the function f such that f'$u = v$ is a choice function for x.

5. $\vdash$ AC $\Rightarrow$ Zorn. Let y partially order a non-empty set x such that every y-chain in x has an upper bound in x. By AC, there is a choice function f for x. Let b be any element of x. By transfinite induction (Proposition 4.14(a)), there is a function F such that F'$\emptyset = b$ and, for any $\alpha >_o \emptyset$, F'α is f'u, where u is the set of y-upper bounds v in x of F"α such that $v \notin F$"α. Let β be the least ordinal such that the set of y-upper bounds in x of F"β that are not in F"β is empty. (There must be such an ordinal. Otherwise, F would be a one–one function with domain On and range a subset of x, which, by the replacement axiom R, would imply that On is a set.) Let $g = \beta\downarrow F$. Then it is easy to check that g is one-one and, if $\alpha <_o \gamma <_o \beta$, $\langle g$'α, g'$\gamma\rangle \in y$. Hence, g"β is a y-chain in x; by hypothesis, there is a y-upper bound w of g"β. Since the set of y-upper bounds of F"$\beta (= g$"$\beta)$ that are not in g"β is empty, $w \in g$"β and w is the only y-upper bound of g"β (because a set can contain at most one of its y-upper bounds). Hence, w is a y-maximal element. (If $\langle w, z\rangle \in y$ and $z \in x$, then z is a y-upper bound of g"β, which is impossible.)

6. $\vdash$ Zorn $\Rightarrow$ WO. Given a set z, let X be the class of all one–one functions with domain an ordinal and range a subset of z. By Hartogs' theorem, X is a set. Clearly, $\emptyset \in X$. X is partially ordered by the proper inclusion relation $\subset$. Given any chain of functions in X, of any two, one is an extension of the other. Hence, the union of all the functions in the chain is also a one–one function from an ordinal into z, which is a $\subset$-upper bound of the chain. Hence, by Zorn, X has a maximal element g, which is a one–one function from an ordinal α into z. Assume $z - g``\alpha \neq \emptyset$ and let $b \in z - g``\alpha$. Let $f = g \cup \{<\alpha, \beta>\}$. Then $f \in X$ and $g \subset f$, contradicting the maximality of g. So, $g``\alpha = z$. Thus, $\alpha \underset{g}{\cong} z$. By means of g, we can transfer the well-ordering E_α of α to a well-ordering of z.

Exercises

4.73 Show that each of the following is equivalent to the axiom of choice.

(a) Any set x is equinumerous with some ordinal.

(b) *Special case of Zorn's lemma.* If x is a non-empty set and if the union of each non-empty $\subset$-chain in x is also in x, then x has a $\subset$-maximal element.

(c) *Hausdorff maximal principle.* If x is a set, then every $\subset$-chain in x is a subset of some maximal $\subset$-chain in x.

(d) *Teichmüller–Tukey Lemma.* Any set of finite character has a $\subset$-maximal element. (A non-empty set x is said to be of *finite character* if and only if: (i) every finite subset of an element of x is also an element of x; and (ii) if every finite subset of a set y is a member of x, then $y \in x$.)

(e) $(\forall x)(\mathrm{Rel}(x) \Rightarrow (\exists y)(\mathrm{Fnc}(y) \wedge \mathscr{D}(x) = \mathscr{D}(y) \wedge y \subseteq x))$

(f) For any non-empty sets x and y, either there is a function with domain x and range y or there is a function with domain y and range x.

4.74 Show that the following finite axiom of choice is provable in NBG: if x is a finite set of nonempty disjoint sets, then there is a choice set y for x. [*Hint*: Assume $x \cong \alpha$ where $\alpha \in \omega$. Use induction on α.]

PROPOSITION 4.43

The following are consequences of the axiom of choice.

(a) Any infinite set has a denumerable subset.

(b) An infinite set is Dedekind-infinite.

(c) If x is a denumberable set whose elements are denumerable sets, then $\bigcup x$ is denumerable.

Proof

Assume AC.

(a) Let x be an infinite set. By Exercise 4.73(a), x is equinumerous with some ordinal α. Since x is infinite, so is α. Hence, $\omega \leqslant_0 \alpha$; therefore, ω is equinumerous with some subset of x.

(b) The proof is by part (a) and Exercise 4.64(c).

(c) Assume x is a denumerable set of denumerable sets. Let f be a function assigning to each u in x the set of all one–one correspondences between u and ω. Let z be the union of the range of f. Then, by AC applied to z, there is a function g such that $g`v \in v$ for each non-empty $v \subseteq z$. In particular, if $u \in X$, then $g`(f`u)$ is a one–one correspondence between $u$ and $\omega$. Let $h$ be a one–one correspondence between $\omega$ and $x$. Define a function $F$ on $\bigcup x$ as follows: let $y \in \bigcup x$ and let $n$ be the smallest element of $\omega$ such that $y \in h`n$. Now, $h`n \in x$; so, $g`(f`(h`n))$ is a one–one correspondence between $h`n$ and $\omega$. Define $F`y = \langle n, (g`(f`(h`n)))`y\rangle$. Then F is a one–one function with domain $\bigcup x$ and range a subset of $\omega \times \omega$. Hence, $\bigcup x \preccurlyeq \omega \times \omega$. But $\omega \times \omega \cong \omega$ and, therefore, $\bigcup x \preccurlyeq \omega$. If $v \in x$, then $v \subseteq \bigcup x$ and $v \cong \omega$. Hence, $\omega \preccurlyeq \bigcup x$. By Bernstein's theorem, $\bigcup x \cong \omega$.

Exercises

4.75 If x is a set, the Cartesian product $\Pi_{u \in x} u$ is the set of functions f with domain x such that $f`u \in u$ for all $u \in x$. Show that AC is equivalent to the proposition that the Cartesian product of any set x of non-empty sets is also non-empty.

4.76 Show that AC implies that any partial ordering of a set x is included in a total ordering of x.

4.77 Prove that the following is a consequence of AC: for any ordinal α, if x is a set such that $x \preccurlyeq \omega_\alpha$ and such that $(\forall u)(u \in x \Rightarrow u \preccurlyeq \omega_\alpha)$, then $\bigcup x \preccurlyeq \omega_\alpha$. [*Hint*: The proof is like that of Proposition 4.43(c).]

4.78 (a) Prove $y \preccurlyeq x \Rightarrow (\exists f)(\mathrm{Fnc}(f) \wedge \mathscr{D}(f) = x \wedge \mathscr{R}(f) = y)$.

(b) Prove that AC implies the converse of part (a).

4.79[D] (a) Prove $(u +_c v)^2 \cong u^2 +_c (2 \times (u \times v)) +_c v^2$.

(b) Assume y is a well-ordered set such that $x \times y \cong x +_c y$ and $\neg(y \preccurlyeq x)$. Prove that $x \preccurlyeq y$.

(c) Assume $y \cong y \times y$ for all infinite sets y. Prove that, if $\mathrm{Inf}(x)$ and $z = \mathscr{H}`x$, then $x \times z \cong x +_c z$.

(d) Prove that AC is equivalent to $(\forall y)(\mathrm{Inf}(y) \Rightarrow y \cong y \times y)$ (Tarski, 1923).

A stronger form of the axiom of choice is the following sentence (UCF): $(\exists X)\,(\mathrm{Fnc}(X) \wedge (\forall u)(u \neq \emptyset \Rightarrow X`u \in u))$. (There is a *universal choice function* – that is, a function that assigns to every non-empty set u an element of u.)

UCF obviously implies AC, but W.B. Easton proved in 1964 that UCF is not provable from AC if NBG is consistent. However, Felgner (1971b) proved that, for any sentence $\mathscr{B}$ in which all quantifiers are restricted to sets, if $\mathscr{B}$ is provable from NBG+(UCF), then $\mathscr{B}$ is provable in NBG+(AC). (See Felgner (1976) for a thorough treatment of the relations between UCF and AC.)

The theory of cardinal numbers can be simplified if we assume AC; for AC implies that every set is equinumerous with some ordinal and, therefore, that every set x is equinumerous with a unique initial ordinal, which can be designated as the *cardinal number* of x. Thus, the cardinal numbers would be identified with the initial ordinals. To conform with the standard notation for ordinals, we let $\aleph_\alpha$ stand for ω_α. Proposition 4.40 and Corollary 4.41 establish some of the basic properties of addition and multiplication of cardinal numbers.

The status of the axiom of choice has become less controversial in recent years. To most mathematicians it seems quite plausible, and it has so many important applications in practically all branches of mathematics that not to accept it would seem to be a wilful hobbling of the practising mathematician. We shall discuss its consistency and independence later in this section.

Another hypothesis that has been proposed as a basic principle of set theory is the so-called *regularity axiom* (Reg):

$$(\forall X)(X \neq \emptyset \Rightarrow (\exists y)(y \in X \wedge y \cap X = \emptyset))$$

(Every non-empty class X contains a member that is disjoint from X.)

PROPOSITION 4.44

(a) The regularity axiom implies the *Fundierungsaxiom*:

$$\neg(\exists f)\mathrm{Fnc}(f) \wedge \mathscr{D}(f) = \omega \wedge (\forall u)(u \in \omega \Rightarrow f`(u') \in f`u))$$

that is, there is no infinitely descending $\in$-sequence $x_0 \ni x_1 \ni x_2 \ni \ldots$

(b) If we assume AC, then the Fundierungsaxiom implies the regularity axiom.

(c) The regularity axiom implies the non-existence of finite $\in$-cycles that is, of functions f on a non-zero finite ordinal α such that $f`\emptyset \in f`1 \in \ldots \in f`\alpha \in f`\emptyset$. In particular, it implies that there is no set y such that $y \in y$.

Proof

(a) Assume $\mathrm{Fnc}(f) \wedge \mathscr{D}(f) = \omega \wedge (\forall u)(u \in \omega \Rightarrow f`(u') \in f`u)$. Let $z = f``\omega$. By (Reg), there is some element y in z such that $y \cap z = \emptyset$. Since $y \in z$, there

is a finite ordinal α such that $y = f`\alpha$. Then $f`(\alpha') \in y \cap z$, contradicting $y \cap z = \emptyset$.

(b) First, we define the *transitive closure* $\mathrm{TC}(u)$ of a set u. Intuitively, we want $\mathrm{TC}(u)$ to be the smallest transitive set that contains u. Define by induction a function g on ω such that $g`\emptyset = \{u\}$ and $g`(\alpha') = \bigcup(g`\alpha)$ for each $\alpha$ in $\omega$. Thus, $g`1 = u, g`2 = \bigcup u, g`3 = \bigcup(\bigcup u)$, and so on. Let $\mathrm{TC}(u) = \bigcup(g``\omega)$ be called the transitive closure of u. For any u, TC(u) is transitive; that is, $(\forall v)(v \in \mathrm{TC}(u) \Rightarrow v \subseteq \mathrm{TC}(u))$. Now, assume AC and the Fundierungsaxiom; also, assume $X \neq \emptyset$ but there is no y in X such that $y \cap X = \emptyset$. Let b be some element of X; hence, $b \cap X \neq \emptyset$. Let $c = \mathrm{TC}(b) \cap X$. By AC, let h be a choice function for c. Define a function f on ω such that $f`\emptyset = b$ and, for any $\alpha$ in $\omega, f`(\alpha') = h`((f`\alpha) \cap X)$. It follows easily that, for each α in $\omega, f`(\alpha') \in f`\alpha$, contradicting the Fundierungsaxiom. (The proof can be summarized as follows: we start with an element b of X; then, using h, we pick an element $f`1$ in $b \cap X$; since, by assumption, $f`1$ and X cannot be disjoint, we pick an element $f`2$ in $f`1 \cap X$, and so on.)

(c) Assume given a finite $\in$-cycle: $f`\emptyset \in f`1 \in \ldots \in f`n \in f`\emptyset$. Let X be the range of f: $\{f`\emptyset, f`1, \ldots, f`n\}$. By (Reg), there is some $f`j$ in X such that $f`j \cap X = \emptyset$. But each element of X has an element in common with X[†].

Exercises

4.80 If z is a transitive set such that $u \in z$, prove that $\mathrm{TC}(u) \subseteq z$.

4.81 By the *principle of dependent choices* (PDC) we mean the following: if r is a non-empty relation whose range is a subset of its domain, then there is a function f: $\omega \to \mathscr{D}(r)$ such that $(\forall u)(u \in \omega \Rightarrow \langle f`u, f`(u')\rangle \in r)$ (Mostowski, 1948).

(a) Prove $\vdash$ AC $\Rightarrow$ PDC.

(b) Show that PDC implies the denumerable axiom of choice (DAC):

$$\mathrm{Den}(x) \wedge (\forall u)(u \in x \Rightarrow u \neq \emptyset) \Rightarrow (\exists f)(f : x \to \bigcup x \wedge (\forall u)(u \in x \Rightarrow f`u \in u))$$

(c) Prove $\vdash$ PDC $\Rightarrow (\forall x)(\mathrm{Inf}(x) \Rightarrow \omega \preccurlyeq x)$ (Hence, by Exercise 4.64(c), PDC implies that a set is infinite if and only if it is Dedekind-infinite.)

(d) Prove that the conjunction of PDC and the Fundierungsaxiom implies (Reg).

Let us define by transfinite induction the following function Ψ with domain On:

[†]The use of AC in deriving (Reg) from the Fundierungsaxiom is necessary. Mendelson (1958) proved that, if NBG is consistent and if we add the Fundierungsaxiom as an axiom, then (Reg) is not provable in this enlarged theory.

$$\Psi`\emptyset = \emptyset$$
$$\Psi`(\alpha') = \mathscr{P}(\Psi`\alpha)$$
$$\mathrm{Lim}(\lambda) \Rightarrow \Psi`\lambda = \bigcup_{\beta <_o \lambda} \Psi`\beta$$

Let H stand for $\bigcup(\Psi``On)$, that is, H consists of all members of sets of the form $\Psi`\alpha$. Let $H_\beta$ stand for $\Psi`(\beta')$. Thus, $H_\beta = \mathscr{P}(\Psi`\beta)$ and $H_{\beta'} = \mathscr{P}(\Psi`(\beta')) = \mathscr{P}(H_\beta)$. In particular, $H_\emptyset = \mathscr{P}(\Psi`\emptyset) = \mathscr{P}(\emptyset) = \{\emptyset\}$, $H_1 = \mathscr{P}(H_\emptyset) = \mathscr{P}(\{\emptyset\}) = \{\emptyset, \{\emptyset\}\}$, and $H_2 = \mathscr{P}(H_1) = \mathscr{P}(\{\emptyset, \{\emptyset\}\}) = \{\emptyset, \{\emptyset\}, \{\{\emptyset\}\}, \{\emptyset, \{\emptyset\}\}\}$.

Define a function ρ on H such that, for any x in H, $\rho`x$ is the least ordinal $\alpha$ such that $x \in \Psi`\alpha$. $\rho`x$ is called the *rank* of $x$. Observe that $\rho`x$ must be a successor ordinal. (In fact, there are no sets of rank $\emptyset$, since $\Psi`\emptyset = \emptyset$. If $\lambda$ is a limit ordinal, every set in $\Psi`\lambda$ already was a member of $\Psi`\beta$ for some $\beta <_o \lambda$.) As examples, note that $\rho`\emptyset = 1, \rho`\{\emptyset\} = 1, \rho`\{\emptyset, \{\emptyset\}\} = 2$, and $\rho`\{\{\emptyset\}\} = 2$.

Exercise 4.82. Prove that the following are theorems of NBG.

(a) $(\forall\alpha)\mathrm{Trans}(\Psi`\alpha)$
(b) $\mathrm{Trans}(H)$
(c) $(\forall\alpha)(\Psi`\alpha \subseteq \Psi`(\alpha'))$
(d) $(\forall\alpha)(\forall\beta)(\alpha <_o \beta \Rightarrow \Psi`\alpha \subseteq \Psi`\beta)$
(e) $On \subseteq H$
(f) $(\forall\alpha)(\rho`\alpha = \alpha')$
(g) $(\forall u)(\forall v)(u \in H \wedge v \in H \wedge u \in v \Rightarrow \rho`u <_o \rho`v)$
(h) $(\forall u)(u \subseteq H \Rightarrow u \in H)$

PROPOSITION 4.45

The regularity axiom is equivalent to the assertion that $V = H$, that is, that every set is a member of H.

Proof

(a) Assume $V = H$. Let $X \neq \emptyset$. Let α be the least of the ranks of all the members of X, and let b be an element of X such that $\rho`b = \alpha$. Then $b \cap X = \emptyset$; for, if $u \in b \cap X$, then, by Exercise 4.82(g), $\rho`u \in \rho`b = \alpha$, contradicting the minimality of α.
(b) Assume (Reg). Assume $V \neq H$. Then $V - H \neq \emptyset$. By (Reg), there is some y in $V - H$ such that $y \cap (V - H) = \emptyset$. Hence, $y \subseteq H$ and so, by Exercise 4.82(h), $y \in H$, contradicting $y \in V - H$.

Exercises

4.83 Show that (Reg) is equivalent to the special case: $(\forall x)(x \neq \emptyset \Rightarrow (\exists y)(y \in x \wedge y \cap x = \emptyset))$.

4.84 Show that, if we assume (Reg), then $\mathrm{Ord}(X)$ is equivalent to $\mathrm{Trans}(X) \wedge \mathrm{E\ Con}\ X$, that is, to the wf

$$(\forall u)(u \in X \Rightarrow u \subseteq X) \wedge (\forall u)(\forall v)(u \in X \wedge v \in X \wedge u \neq v \Rightarrow u \in v \vee v \in u)$$

Thus, with the regularity axiom, a much simpler definition of the notion of ordinal class is available, a definition in which all quantifiers are restricted to sets.

4.85 Show that (Reg) implies that every non-empty transitive class contains $\emptyset$

Proposition 4.45 certainly increases the attractiveness of adding (Reg) as a new axiom to NBG. The proposition $V = H$ asserts that every set can be obtained by starting with $\emptyset$ and applying the power set and union operations any transfinite number of times. The assumption that this is so is called the *iterative conception of set*. Many set theorists now regard this conception as the best available formalization of our intuitive picture of the universe of sets.†

By Exercise 4.84, the regularity axiom would also simplify the definition of ordinal numbers. In addition, we can develop the theory of cardinal numbers on the basis of the regularity axiom; namely, just define the cardinal number of a set x to be the set of all those y of lowest rank such that $y \cong x$. This would satisfy the basic requirement of a theory of cardinal numbers, the existence of a function Card whose domain is V and such that $(\forall x)(\forall y)(\mathrm{Card}`x = \mathrm{Card}`y \Leftrightarrow x \cong y)$.

There is no unanimity among mathematicians about whether we have sufficient grounds for adding (Reg) as a new axiom, for, although it has great simplifying power, it does not have the immediate plausibility that even the axiom of choice has, nor has it had any mathematical applications. Nevertheless, it is now often taken without explicit mention to be one of the axioms.

The class H determines an *inner model* of NBG in the following sense. For any wf $\mathscr{B}$ (written in unabbreviated notation), let $\mathrm{Rel}_H(\mathscr{B})$ be the wf obtained from $\mathscr{B}$ by replacing every subformula $(\forall X)\mathscr{C}(X)$ by $(\forall X)(X \subseteq H \Rightarrow \mathscr{C}(X))$ (in making the replacements we start with the in-

†The iterative conception seems to presuppose that we understand the power set and union operations and that ordinal numbers (or something essentially equivalent to them) are available for carrying out the transfinite iteration of the power set and union operations.

nermost subformulas) and then, if $\mathscr{B}$ contains free variables. $Y_1, \ldots, Y_n$, prefixing $(Y_1 \subseteq H \wedge Y_2 \subseteq H \wedge \ldots \wedge Y_n \subseteq H) \Rightarrow$.

In other words, in forming $\mathrm{Rel}_H(\mathscr{B})$, we interpret 'class' as 'subclass of H'. Since $\mathrm{M}(X)$ stands for $(\exists Y)(X \in Y)$, $\mathrm{Rel}_H(\mathrm{M}(X))$ is $(\exists Y)(Y \subseteq H \wedge X \in Y)$, which is equivalent to $X \in H$; thus, the 'sets' of the model are the elements of H. Hence, $\mathrm{Rel}_H((\forall x)\mathscr{B})$ is equivalent to $(\forall x)(x \in H \Rightarrow \mathscr{B}^{\#})$, where $\mathscr{B}^{\#}$ is $\mathrm{Rel}_H(\mathscr{B})$. Note also that $\vdash X \subseteq H \wedge Y \subseteq H \Rightarrow [\mathrm{Rel}_H(X = Y) \Leftrightarrow X = Y]$. Then it turns out that, for any theorem $\mathscr{B}$ of NBG, $\mathrm{Rel}_H(\mathscr{B})$ is also a theorem of NBG.

Exercises

4.86 Verify that, for each axiom $\mathscr{B}$ of NBG, $\vdash_{\mathrm{NBG}} \mathrm{Rel}_H(\mathscr{B})$. If we adopt a semantic approach, one need only show that, if $\mathscr{M}$ is a model for NBG, in the usual sense of 'model', then the objects X of $\mathscr{M}$ that satisfy the wf $X \subseteq H$ also form a model for NBG. In addition, one can verify that (Reg) holds in this model; this is essentially just part (a) of Proposition 4.45. A direct consequence of this fact is that, if NBG is consistent, then so is the theory obtained by adding (Reg) as a new axiom. That (Reg) is independent of NBG (that is, cannot be proved in NBG) can be shown by means of a model that is somewhat more complex than the one given above for the consistency proof (see Bernays, 1937–1954, part VII). Thus, we can consistently add either (Reg) or its negation to NBG, if NBG is consistent. Practically the same arguments show the independence and consistency of (Reg) with respect to NBG + (AC).

4.87 Consider the model whose domain is H_α and whose interpretation of $\in$ is E_{H_α}, the membership relation restricted to H_α. Notice that the 'sets' of this model are the sets of rank $\leqslant_0 \alpha$ and the 'proper classes' are the sets of rank α'. Show that the model H_α satisfies all axioms of NBG (except possibly the axioms of infinity and replacement) if and only if $\mathrm{Lim}(\alpha)$. Prove also that H_α satisfies the axiom of infinity if and only if $\alpha >_0 \omega$.

4.88 Show that the axiom of infinity is not provable from the other axioms of NBG, if the latter form a consistent theory.

4.89 Show that the replacement axiom (R) is not provable from the other axioms (T, P, N, (B1) (B7), U, W, S) if these latter form a consistent theory.

4.90 An ordinal α such that H_α is a model for NBG is called *inaccessible*. Since NBG has only a finite number of proper axioms, the assertion that α is inaccessible can be expressed by the conjunction of the relativization to H_α of the proper axioms of NBG. Show that the existence of inaccessible ordinals is not provable in NBG if NBG is consistent. (Compare Shepherdson (1951–53), Montague and Vaught (1959), and, for related results, Bernays (1961) and Levy (1960).) Inaccessible ordinals have been shown to have connections with problems in measure theory and algebra (see Ulam, 1930;

Zeeman, 1955; Erdös and Tarski, 1961).† The consistency of the theory obtained from NBG by adding an axiom asserting the existence of an inaccessible ordinal is still an open question. More about inaccessible ordinals may be found in Exercise 4.91.

The axiom of choice turns out to be consistent and independent with respect to the theory NBG + (Reg). More precisely, if NBG is consistent, AC is an undecidable sentence of the theory NBG+ (Reg). In fact, Gödel (1938; 1939; 1940) showed that, if NBG is consistent, then the theory NBG + (AC) + (Reg) + (GCH) is also consistent, where (GCH) stands for the *generalized continuum hypothesis*:

$$(\forall x)(\mathrm{Inf}(x) \Rightarrow \neg(\exists y)(x \prec y \wedge y \prec \mathscr{P}(x)))$$

(Our statement of Gödel's result is a bit redundant, since $\vdash_{\mathrm{NBG}}$ (GCH) $\Rightarrow$ (AC) has been proved by Sierpinski (1947) and Specker (1954). This result will be proved below.) The unprovability of AC from NBG + (Reg), if NBG is consistent, has been proved by P.J. Cohen (1963–64), who also has shown the independence of the special *continuum hypothesis*, $2^{\omega} \cong \omega_1$, in the theory NBG + (AC) + (Reg). Expositions of the work of Cohen and its further development can be found in Cohen (1966) and Shoenfield (1971b), as well as in Rosser (1969) and Felgner (1971a). For a thorough treatment of these results and other independence proofs in set theory, Jech (1978) and Kunen (1980) should be consulted.

We shall present here a modified form of the proof in Cohen (1966) of Sierpinski's theorem that GCH implies AC.

DEFINITION

For any set v, let $\mathscr{P}^0(v) = v, \mathscr{P}^1(v) = \mathscr{P}(v), \mathscr{P}^2(v) = \mathscr{P}(\mathscr{P}(v)), \ldots, \mathscr{P}^{k+1}(v) = \mathscr{P}(\mathscr{P}^k(v))$ for all k in ω.

LEMMA 4.46

If $\omega \preccurlyeq v$, then $\mathscr{P}^k(v) +_{\mathrm{c}} \mathscr{P}^k(v) \cong \mathscr{P}^k(v)$ for all $k \geqslant_{\mathrm{o}} 1$.

†Inaccessible ordinals are involved also with attempts to provide a suitable set-theoretic foundation for category theory (see MacLane, 1971; Gabriel, 1962; Sonner, 1962; Kruse, 1966; Isbell, 1966).

Proof

Remember that $\mathscr{P}(x) \cong 2^x$ (see Exercise 4.40). From $\omega \preccurlyeq v$ we obtain $\omega \preccurlyeq \mathscr{P}^k(v)$ for all k in ω. Hence, $\mathscr{P}^k(v) +_c 1 \cong \mathscr{P}^k(v)$ for all k in ω, by Exercise 4.64(g). Now, for any $k \geq_o 1$,

$$\mathscr{P}^k(v) +_c \mathscr{P}^k(v) = \mathscr{P}^k(v) \times 2 = \mathscr{P}(\mathscr{P}^{k-1}(v)) \times 2 \cong 2^{\mathscr{P}^{k-1}(v)} \times 2$$
$$\cong 2^{\mathscr{P}^{k-1}(v)} \times 2^1 \cong 2^{\mathscr{P}^{k-1}(v)+_c 1} \cong 2^{\mathscr{P}^{k-1}(v)} \cong \mathscr{P}(\mathscr{P}^{k-1}(v)) = \mathscr{P}^k(v)$$

LEMMA 4.47

If $y +_c x \cong \mathscr{P}(x +_c x)$, then $\mathscr{P}(x) \preccurlyeq y$.

Proof

Notice that $\mathscr{P}(x +_c x) \cong 2^{x+_c x} \cong 2^x \times 2^x \cong \mathscr{P}(x) \times \mathscr{P}(x)$. Let $y^* = y \times \{\emptyset\}$ and $x^* = x \times \{1\}$. Since $y +_c x \cong \mathscr{P}(x +_c x) \cong \mathscr{P}(x) \times \mathscr{P}(x)$, there is a function f such that $y^* \cup x^* \overset{f}{\cong} \mathscr{P}(x) \times \mathscr{P}(x)$. Let h be the function that takes each u in x^* into the first component of the pair $f`u$. Thus, $h: x^* \Rightarrow \mathscr{P}(x)$. By Proposition 4.25(a), there must exist c in $\mathscr{P}(x) - h``(x^*)$. Then, for all z in $\mathscr{P}(x)$, there exists a unique v in y^* such that $f`v = \langle c, z\rangle$. This determines a one-one function from $\mathscr{P}(x)$ into y. Hence, $\mathscr{P}(x) \preccurlyeq y$.

PROPOSITION 4.48

Assume GCH.

(a) If u cannot be well-ordered and $u +_c u \cong u$ and β is an ordinal such that $\beta \preccurlyeq 2^u$, then $\beta \preccurlyeq u$.
(b) The axiom of choice holds.

Proof

(a) Notice that $u +_c u \cong u$ implies $1 +_c u \cong u$, by Exercise 4.71(b). Therefore, by Exercise 4.55(i), $2^u +_c u \cong 2^u$. Now, $u \preccurlyeq \beta +_c u \preccurlyeq 2^u +_c u \cong 2^u$. By GCH, either (i) $u \cong \beta +_c u$ or (ii) $\beta +_c u \cong 2^u$. If (ii) holds, $\beta +_c u \cong 2^u +_c u \cong \mathscr{P}(u +_c u)$. Hence, by Lemma 4.47, $\mathscr{P}(u) \preccurlyeq \beta$ and, therefore, $u \preccurlyeq \beta$. Then, since u would be equinumerous with a subset of an ordinal, u could be well-ordered, contradicting our assumption. Hence, (i) must hold. But then, $\beta \preccurlyeq \beta +_c u \cong u$.

(b) We shall prove AC by proving the equivalent sentence (WO) asserting that every set can be well-ordered. To that end, consider any set x and

assume, for the sake of contradiction, that x cannot be well-ordered. Let $v = 2^{x \cup \omega}$. Then $\omega \preccurlyeq x \cup \omega \preccurlyeq v$. Hence, by Lemma 4.46, $\mathscr{P}^k(v) +_c \mathscr{P}^k(v) \cong \mathscr{P}^k(v)$ for all $k \geqslant_o 1$. Also, since $x \preccurlyeq x \cup \omega \preccurlyeq v \prec \mathscr{P}(v) \prec \mathscr{P}(\mathscr{P}(v)) \prec \ldots$, and x cannot be well-ordered, each $\mathscr{P}^k(v)$ cannot be well-ordered, for $k \geqslant_o 0$. Let $\beta = \mathscr{H}'v$. We know that $\beta \preccurlyeq \mathscr{P}^4(v)$ by Corollary 4.32. Hence, by part (a), with $u = \mathscr{P}^3(v)$, we obtain $\beta \preccurlyeq \mathscr{P}^3(v)$. Using part (a) twice more (successively with $u = \mathscr{P}^2(v)$ and $u = \mathscr{P}(v)$), we obtain $\mathscr{H}'v = \beta \preccurlyeq v$. But this contradicts the definition of $\mathscr{H}'v$ as the least ordinal not equinumerous with a subset of v.

Exercise

4.91 An *α-sequence* is defined to be a function f whose domain is α. If the range of f consists of ordinals, then f is called an *ordinal α-sequence* and, if, in addition, $\beta <_o \gamma <_o \alpha$ implies $f'\beta <_o f'\gamma$, then f is called an *increasing ordinal α-sequence*. By Proposition 4.12, if f is an increasing ordinal α-sequence, then $\bigcup(f''\alpha)$ is the least upper bound of the range of f. An ordinal δ is said to be *regular* if, for any increasing ordinal α-sequence such that $\alpha <_o \delta$ and the ordinals in the range of f are all $<_o \delta$, $\bigcup(f''\alpha) +_o 1 <_o \delta$. Non-regular ordinals are called *singular* ordinals.

(a) Which finite ordinals are regular?

(b) Show that ω_0 is regular and ω_ω is singular

(c) Prove that every regular ordinal is an initial ordinal.

(d) Assuming the axiom of choice (AC), prove that every ordinal of the form $\omega_{\gamma+_o 1}$ is regular.

(e) If ω_α is regular and Lim(α), prove that $\omega_\alpha = \alpha$. (A regular ordinal ω_α such that Lim(α) is called a *weakly inaccessible ordinal.*)

(f) Show that, if ω_α has the property that $\gamma <_o \omega_\alpha$ implies $\mathscr{P}(\gamma) \prec \omega_\alpha$, then Lim($\alpha$). The converse is implied by the generalized continuum hypothesis. A regular ordinal ω_α such that $\alpha >_o \emptyset$ and such that $\gamma <_o \omega_\alpha$ implies $\mathscr{P}(\gamma) \prec \omega_\alpha$, is called *strongly inaccessible*. Thus, every strongly inaccessible ordinal is weakly inaccessible and, if (GCH) holds, the strongly inaccessible ordinals coincide with the weakly inaccessible ordinals.

(g) (Sheperdson 1951–53; Montague and Vaught, 1959) (i) If γ is inaccessible (i.e., if H_γ is a model of NBG), prove that γ is weakly inaccessible. (ii)^D In the theory NBG + (AC), show that γ is inaccessible if and only if γ is strongly inaccessible.

(h) If NBG is consistent, then in the theory NBG + (AC) + (GCH), show that it is impossible to prove the existence of weakly inaccessible ordinals.

4.6 OTHER AXIOMATIZATIONS OF SET THEORY

We have chosen to develop set theory on the basis of NBG because it is relatively simple and convenient for the practising mathematician. There are, of course, many other varieties of axiomatic set theory, of which we will now make a brief survey.

Morse–Kelley (MK)

Strengthening NBG, we can replace axioms (B1)–(B7) by the axiom schema:

$$(\square) \qquad (\exists Y)(\forall x)(x \in Y \Leftrightarrow \mathscr{B}(x))$$

where $\mathscr{B}(x)$ is *any* wf (not necessarily predicative) of NBG and Y is not free in $\mathscr{B}(x)$. The new theory MK, called *Morse–Kelley set theory*, became well-known through its appearance as an appendix in a book on general topology by Kelley (1955). The basic idea was proposed independently by Mostowski, Quine, and Morse (whose rather unorthodox system may be found in Morse (1965)). Axioms (B1)–(B7) follow easily from ($\square$) and, therefore, NBG is a subtheory of MK. Mostowski (1951a) showed that, if NBG is consistent, then MK is really stronger than NBG. He did this by constructing a 'truth definition' in MK on the basis of which he proved $\vdash_{\mathrm{MK}} \mathscr{C}on_{\mathrm{NBG}}$, where $\mathscr{C}on_{\mathrm{NBG}}$ is a standard arithmetic sentence asserting the consistency of NBG. On the other hand, by Gödel's second theorem, $\mathscr{C}on_{\mathrm{NBG}}$ is not provable in NBG if the latter is consistent.

The simplicity and power of schema ($\square$) make MK very suitable for use by mathematicians who are not interested in the subtleties of axiomatic set theory. But this very strength makes the consistency of MK a riskier gamble. However, if we add to NBG + (AC) the axiom (In) asserting the existence of a strongly inaccessible ordinal θ, then H_θ is a model of MK. Hence, MK involves no more risk than NBG + (AC) + (In).

There are several textbooks that develop axiomatic set theory on the basis of MK (Rubin, 1967; Monk, 1980; Chuquai, 1981). Some of Cohen's independence results have been extended to MK by Chuquai (1972).

Exercises

4.92 Prove that axioms (B1)–(B7) are theorems of MK.
4.93 Verify that, if θ is a strongly inaccessible ordinal, then H_θ is a model of MK.

Zermelo–Fraenkel (ZF)

The earliest axiom system for set theory was devised by Zermelo (1908). The objects of the theory are thought of intuitively as *sets*, not the *classes* of NBG or MK. Zermelo's theory Z can be formulated in a language that contains only one predicate letter $\in$. Equality is defined extensionally: $x = y$ stands for $(\forall z)(z \in x \Leftrightarrow z \in y)$. The proper axioms are:

T: $x = y \Rightarrow (x \in z \Leftrightarrow y \in z)$ (*substitutivity of* $=$)
P: $(\exists z)(\forall u)(u \in z \Leftrightarrow u = x \vee u = y)$ (*pairing*)
N: $(\exists x)(\forall y)(y \notin x)$ (*null set*)
U: $(\exists y)(\forall u)(u \in y \Leftrightarrow (\exists v)(u \in v \wedge v \in x))$ (*sum set*)
W: $(\exists y)(\forall u)(u \in y \Leftrightarrow u \subseteq x)$ (*power set*)
S^*: $(\exists y)(\forall u)(u \in y \Leftrightarrow (u \in x \wedge \mathscr{B}(u)))$, where $\mathscr{B}(u)$ is any wf not containing y free (*selection*)
I: $(\exists x)(\emptyset \in x \wedge (\forall z)(z \in x \Rightarrow z \cup \{z\} \in x))$ (*infinity*)

Here we have assumed the same definitions of $\subseteq, \emptyset, \cup$ and $\{u\}$ as in NBG.

Zermelo's intention was to build up mathematics by starting with a few simple sets ($\emptyset$ and ω) and then constructing further sets by various well-defined operations (such as formation of pairs, unions and power sets). In fact, a good deal of mathematics can be built up within Z. However, Fraenkel (1922a) observed that Z was too weak for a full development of mathematics. For example, for each finite ordinal n, the ordinal $\omega +_0 n$ can be shown to exist, but the set A of all such ordinals cannot be proved to exist, and, therefore, $\omega +_0 \omega$, the least upper bound of A, cannot be shown to exist. Fraenkel proposed a way of overcoming such difficulties, but his idea could not be clearly expressed in the language of Z. However, Skolem (1923) was able to recast Fraenkel's idea in the following way: for any wf $\mathscr{B}(x, y)$, let Fun($\mathscr{B}$) stand for $(\forall x)(\forall u)(\forall v)(\mathscr{B}(x, u) \wedge \mathscr{B}(x, v) \Rightarrow u = v)$. Thus, Fun ($\mathscr{B}$) asserts that $\mathscr{B}$ determines a function. Skolem's *axiom schema of replacement* can then be formulated as follows:

$$(\mathrm{R}^{\#})\ \mathrm{Fun}(\mathscr{B}) \Rightarrow (\forall w)(\exists z)(\forall v)(v \in z \Leftrightarrow (\exists u)(u \in w \wedge \mathscr{B}(u, v)))$$
for any wf $\mathscr{B}(x, y)$

This is the best approximation that can be found for the replacement axiom R of NBG.

The system $Z + (R^{\#})$ is denoted ZF and is called *Zermelo–Fraenkel set theory*. In recent years, ZF is often assumed to contain a set-theoretic regularity axiom (Reg*): $x \neq \emptyset \Rightarrow (\exists y)(y \in x \wedge y \cap x = \emptyset)$. The reader should always check to see whether (Reg*) is included within ZF. ZF is now the most popular form of axiomatic set theory; most of the modern research in set theory on independence and consistency proofs has been carried out with

respect to ZF. For expositions of ZF, see Krivine (1971), Suppes (1960), Zuckerman (1974), Lévy (1978) and Hrbacek and Jech (1978).

ZF and NBG yield essentially equivalent developments of set theory. Every sentence of ZF is an abbreviation of a sentence of NBG since, in NBG, lower-case variables $x, y, z, \ldots$ serve as restricted set variables. Thus axiom N is an abbreviation of $(\exists x)(\mathrm{M}(x) \wedge (\forall y)(\mathrm{M}(y) \Rightarrow y \notin x))$ in NBG. It is a simple matter to verify that all axioms of ZF are theorems in NBG. Indeed, NBG was originally constructed so that this would be the case. We can conclude that, if NBG is consistent, then so is ZF. In fact, if a contradiction could be derived in ZF, the same proof would yield a contradiction in NBG.

The presence of class variables in NBG seems to make it much more powerful than ZF. At any rate, it is possible to express propositions in NBG that either are impossible to formulate in ZF (such as the universal choice axiom) or are much more unwieldy in ZF (such as transfinite induction theorems). Nevertheless, it is a surprising fact that NBG is no riskier than ZF. An even stronger result can be proved: NBG is a *conservative extension* of ZF in the sense that, for any sentence $\mathscr{B}$ of the language of ZF, if $\vdash_{\mathrm{NBG}} \mathscr{B}$, then $\vdash_{\mathrm{ZF}} \mathscr{B}$ (see Novak (Gal) 1951; Rosser and Wang, 1950; Shoenfield, 1954). This implies that, if ZF is consistent, then NBG is consistent. Thus, NBG is consistent if and only if ZF is consistent, and NBG seems to be no stronger than ZF. However, NBG and ZF do differ with respect to the existence of certain kinds of models (see Montague and Vaught, 1959). Moreover, another important difference is that NBG is finitely axiomatizable, whereas Montague (1961a) showed that ZF (as well as Z) is not finitely axiomatizable. Montague (1961b) proved the stronger result that ZF cannot be obtained by adding a finite number of axioms to Z.

Exercise

4.94 Let $H_\alpha^* = \bigcup H_\alpha$ (see page 281).

(a) Verify that H_α^* consists of all sets of rank less than α.

(b) If α is a limit ordinal $>_{\mathrm{o}} \omega$, show that H_α^* is a model for Z.

(c)$^{\mathrm{D}}$ Find an instance of the axiom schema of replacement ($\mathrm{R}^{\#}$) that is false in $H^*_{\omega +_{\mathrm{o}} \omega}$. [*Hint*: Let $\mathscr{B}(x, y)$ be $x \in \omega \wedge y = \omega +_{\mathrm{o}} x$. Observe that $\omega +_{\mathrm{o}} \omega \notin H^*_{\omega +_{\mathrm{o}} \omega}$ and $\omega +_{\mathrm{o}} \omega = \bigcup\{v \mid (\exists u)(u \in \omega \wedge \mathscr{B}(u, v))\}$.]

(d) Show that, if ZF is consistent, then ZF is a proper extension of Z.

The theory of types (ST)

Russell's paradox is based on the set K of all those sets that are not members of themselves: $K = \{x \mid x \notin x\}$. Clearly, $K \in K$ if and only if $K \notin K$. In NBG

this argument simply shows that K is a proper class, not a set. In ZF the conclusion is just that there is no such set K.

Russell himself chose to find the source of his paradox elsewhere. He maintained that $x \in x$ and $x \notin x$ should be considered 'illegitimate' and 'ungrammatical' formulas and, therefore, that the definition of K makes no sense. However, this alone is not adequate because paradoxes analogous to Russell's can be obtained from slightly more complicated circular properties, like $x \in y \wedge y \in x$.

Exercise

4.95 (a) Derive a Russell-style paradox by using $x \in y \wedge y \in x$.
(b) Use $x \in y_1 \wedge y_1 \in y_2 \wedge \ldots \wedge y_{n-1} \in y_n \wedge y_n \in x$ to obtain a paradox, where $n \geq 1$.

Thus, to avoid paradoxes, one must forbid any kind of indirect circularity. For this purpose, we can think of the universe divided up into types in the following way. Start with a collection W of *non-sets* or *individuals*. The elements of W are said to have *type* 0. Sets whose members are of type 0 are the objects of type 1. Sets whose members are of type1 will be the objects of type 2, and so on.

Our language will have variables of different types. The superscript of a variable will indicate its type. Thus, x^0 is a variable of type 0, y^1 is a variable of type 1, and so on. There are no variables other than type variables. The atomic wfs are of the form $x^n \in y^{n+1}$, where n is one of the natural numbers $0, 1, 2, \ldots$. The rest of the wfs are built up from the atomic wfs by means of logical connectives and quantifiers. Observe that $\neg(x \in x)$ and $\neg(x \in y \wedge y \in x)$ are not wfs.

The equality relation must be defined piecemeal, one definition for each type.

DEFINITION

$x^n = y^n$ for $(\forall z^{n+1})(x^n \in z^{n+1} \Leftrightarrow y^n \in z^{n+1})$ Notice that two objects are defined to be equal if they belong to the same sets of the next higher type. The basic property of equality is provided by the following axiom scheme.

ST1 (EXTENSIONALITY AXIOM)

$$(\forall x^n)(x^n \in y^{n+1} \Leftrightarrow x^n \in z^{n+1}) \Rightarrow y^{n+1} = z^{n+1}$$

This asserts that two sets that have the same members must be equal. On the other hand, observe that the property of having the same members could

not be taken as a general definition of equality because it is not suitable for objects of type 0.

Given any wf $\mathscr{B}(x^n)$, we wish to be able to define a set $\{x^n \mid \mathscr{B}(x^n)\}$.

ST2 (COMPREHENSION AXIOM SCHEME)

For any wf $\mathscr{B}(x^n)$, the following wf is an axiom:

$$(\exists y^{n+1})(\forall x^n)(x^n \in y^{n+1} \Leftrightarrow \mathscr{B}(x^n))$$

Here, y^{n+1} is any variable not free in $\mathscr{B}(x^n)$. If we use the extensionality axiom, then the set y^{n+1} asserted to exist by axiom ST2 is unique and can be denoted by $\{x^n \mid \mathscr{B}(x^n)\}$.

Within this system, we can define the usual set-theoretic notions and operations, as well as the natural numbers, ordinal numbers, cardinal numbers and so on. However, these concepts are not unique but are repeated for each type (or, in some cases, for all but the first few types). For example, the comprehension scheme provides a null set $\Lambda^{n+1} = \{x^n \mid x^n \neq x^n\}$ for each non-zero type. But there is no null set *per se*. The same thing happens for natural numbers. n type theory, the natural numbers are not defined as they are in NBG. Here they are the finite cardinal numbers. For example, the set of natural numbers of type 2 is the intersection of all sets containing $\{\Lambda^1\}$ and closed under the following successor operation: the *successor* $S(y^2)$ of a set y^2 is $\{v^1 \mid (\exists u^1)(\exists z^0)\ (u^1 \in y^2 \wedge z^0 \notin u^1 \wedge v^1 = u^1 \cup \{z^0\})\}$. Then, among the natural numbers of type 2, we have $0 = \{\Lambda^1\}, 1 = S(0), 2 = S(1)$, and so on. Here, the numerals 0, 1, 2, ... should really have a superscript ^{2} to indicate their type, but the superscripts were omitted for the sake of legibility. Note that 0 is the set of all sets of type 1 that contain no elements, 1 is the set of all sets of type 1 that contain one element, 2 is the set of all sets of type 1 that contain two elements, and so on.

This repetition of the same notion in different types makes it somewhat inconvenient for mathematicians to work within a type theory. Moreover, it is easy to show that the existence of an infinite set cannot be proved from the extensionality and comprehension schemas.† To see this, consider the 'model' in which each variable of type n ranges over the sets of rank less than or equal to $n +_0 1$. (There is nothing wrong about assigning overlapping ranges to variables of different types.)

We shall assume an axiom that guarantees the existence of an infinite set. As a preliminary, we shall adopt the usual definition $\{\{x^n\}, \{x^n, y^n\}\}$ of the ordered pair: $\langle x^n, y^n \rangle$, where $\{x^n, y^n\}$ stands for $\{u^n \mid u^n = x^n \vee u^n = y^n\}$.

†This fact seemed to undermine Russell's doctrine of *logicism*, according to which all of mathematics could be reduced to basic axioms that were of an essentially *logical* character. An axiom of infinity could not be thought of as a logical truth.

Notice that $\langle x^n, y^n \rangle$ is of type $n+2$. Hence, a binary relation on a set A, being a set of ordered pairs of elements of A, will have type 2 greater than the type of A. In particular, a binary relation on the universe $V^1 = \{x^0 \mid x^0 = x^0\}$ of all objects of type 0 will be a set of type 3.

ST3 (AXIOM OF INFINITY)

$$(\exists x^3)([(\exists u^0)(\exists v^0)(\langle u^0, v^0 \rangle \in x^3)] \wedge$$
$$(\forall u^0)(\forall v^0)(\forall w^0)(\langle u^0, u^0 \rangle \notin x^3 \wedge [\langle u^0, v^0 \rangle \in x^3 \wedge \langle v^0, w^0 \rangle \in x^3 \Rightarrow$$
$$\langle u^0, w^0 \rangle \in x^3] \wedge [\langle u^0, v^0 \rangle \in x^3 \Rightarrow (\exists z^0)(\langle v^0, z^0 \rangle \in x^3)]))$$

This asserts that there is a non-empty irreflexive, transitive binary relation x^3 on V^1 such that every member of the range of x^3 also belongs to the domain of x^3. Since no such relation exists on a finite set, V^1 must be infinite.

The system based on ST1- ST3 is called the *simple theory of types* and is denoted ST. Because of its somewhat complex notation and the repetition of concepts at all (or, in some cases, almost all) type levels, ST is not generally used as a foundation of mathematics and is not the subject of much contemporary research. Suggestions by Turing (1948) to make type theory more usable have been largely ignored.

With ST we can associate a first-order theory ST*. The non-logical constants of ST* are $\in$ and monadic predicates T_n for each natural number n. We then translate any wf $\mathscr{B}$ of ST into ST* by replacing subformulas $(\forall x^n)\mathscr{C}(x^n)$ by $(\forall x)(T_n(x) \Rightarrow \mathscr{C}(x^n))$ and, finally, if $y^{j_1}, \ldots, y^{j_k}$ are the free variables of $\mathscr{B}$, prefixing to the result $T_{j_1}(y_1) \wedge \ldots \wedge T_{j_k}(y_k) \Rightarrow$ and changing each y^{j_i} into y_i. In a rigorous presentation, we would have to specify clearly that the replacements are made by proceeding from smaller to larger subformulas and that the variables $x, y_1, \ldots, y_k$ are new variables. The axioms of ST* are the translations of the axioms of ST. Any theorem of ST translates into a theorem of ST*.

Exercise

4.96 Exhibit a model of ST* within NBG.

By virtue of Exercise 4.96, NBG (or ZF) is stronger than ST: (1) any theorem of ST can be translated into a corresponding theorem of NBG; and (2) if NBG is consistent, so is ST.†

To provide a type theory that is easier to work with, one can add axioms that impose additional structure on the set V^1 of objects of type 0. For

†A stronger result was proved by John Kemeny (1949) by means of a truth definition within Z: if Z is consistent, so is ST.

example, Peano's axioms for the natural numbers were adopted at level 0 in Gödel's system P, for which he originally proved his famous incompleteness theorem (see Gödel, 1931).

In *Principia Mathematica* (1910–1913), the three-volume work by Alfred North Whitehead and Bertrand Russell, there is a theory of types that is further complicated by an additional hierarchy of *orders*. This hierarchy was introduced so that the comprehension scheme could be suitably restricted in order not to generate an *impredicatively defined* set, that is, a set A defined by a formula in which some quantified variable ranges over a set that turns out to contain the set A itself. Along with the mathematician Henri Poincaré, Whitehead and Russell believed impredicatively defined sets to be the root of all evil. However, such concepts are required in analysis (for example, in the proof that any non-empty set of real numbers that is bounded above has a least upper bound). *Principia Mathematica* had to add the so-called *axiom of reducibility* to overcome the order restrictions imposed on the comprehension scheme. The Whitehead–Russell system without the axiom of reducibility is called *ramified type theory*; it is mathematically weak but is of interest to those who wish an extreme constructivist approach to mathematics. The axiom of reducibility vitiates the effect of the order hierarchy; therefore, it is much simpler to drop the notion of order and the axiom of reducibility. The result is the simple theory of types ST, which we have described above.

In ST, the types are natural numbers. For a smoother presentation, some logicians allow a larger set of types, including types for relations and/or functions defined on objects taken from previously defined types. Such a system may be found in Church (1940).

Principia Mathematica must be read critically; for example, it often overlooks the distinction between a formal theory and its metalanguage. The idea of a simple theory of types goes back to Ramsey (1925) and, independently, to Chwistek (1924–25). Discussions of type theory are found in Andrews (1986), Hatcher (1982) and Quine (1963).

Quine's theories NF and ML

Quine (1937) invented a type theory that was designed to do away with some of the unpleasant aspects of type theory while keeping the essential idea of the comprehension axiom ST2. Quine's theory NF (New Foundations) uses only one kind of variable $x, y, z, \ldots$ and one binary predicate letter $\in$. Equality is defined as in type theory: $x = y$ stands for $(\forall z)(x \in z \Leftrightarrow y \in z)$. The first axiom is familiar:

NF1 (EXTENSIONALITY)

$$(\forall z)(z \in x \Leftrightarrow z \in y) \Rightarrow x = y$$

In order to formulate the comprehension axiom, we introduce the notion of *stratification*. A wf $\mathscr{B}$ is said to be *stratified* if one can assign integers to the variables of $\mathscr{B}$ so that: (1) all occurrences of the same free variable are assigned the same integer; (2) all bound occurrences of a variable that are bound by the same quantifier must be assigned the same integer; and (3) for every subformula $x \in y$ of $\mathscr{B}$, the integer assigned to y is 1 greater than the integer assigned to x.

Examples

1. $(\exists y)(x \in y \wedge y \in z) \vee u \in x$ is stratified by virtue of the assignment indicated below by superscripts:
$$(\exists y^2)(x^1 \in y^2 \wedge y^2 \in z^3) \vee u^0 \in x^1$$
2. $((\exists y)(x \in y)) \wedge (\exists y)(y \in x)$ is stratified as follows:
$$((\exists y^2)(x^1 \in y^2)) \wedge (\exists y^0)(y^0 \in x^1)$$
Notice that the ys in the second conjunct do not have to have the same integers assigned to them as the ys in the first conjunct.
3. $x \in y \vee y \in x$ is not stratified. If x is assigned an integer n, then the first y must be assigned $n+1$ and the second y must be assigned $n-1$, contradicting (1).

NF2 (COMPREHENSION)

For any stratified wf $\mathscr{B}(x)$,

$$(\exists y)(\forall x)(x \in y \Leftrightarrow \mathscr{B}(x))$$

is an axiom. (Here, y is assumed to be the first variable not free in $\mathscr{B}(x)$.)

Although NF2 is an axiom scheme, it turns out that NF is finitely axiomatizable (Hailperin, 1944).

Exercise

4.97 Prove that equality could have been defined as follows: $x = y$ for $(\forall z)(x \in z \Rightarrow y \in z)$ (More precisely, in the presence of NF2, this definition is equivalent to the original one.)

The theory of natural numbers, ordinal numbers and cardinal numbers is developed in much the same way as in type theory, except that there is no longer a multiplicity of similar concepts. There is a unique empty set $\Lambda = \{x \mid x \neq x\}$ and a unique universal set $V = \{x \mid x = x\}$. We can easily

prove $V \in V$, which immediately distinguishes NF from type theory (and from NBG, MK and ZF).

The usual argument for Russell's paradox does not hold in NF, since $x \notin x$ is not stratified. Almost all of standard set theory and mathematics is derivable in NF; this is done in full detail in Rosser (1953). However, NF has some very strange properties. First of all, the usual proof of Cantor's theorem, $A \prec \mathscr{P}(A)$, does not go through in NF; at a key step in the proof, a set that is needed is not available because its defining condition is not stratified. The apparent unavailability of Cantor's theorem has the desirable effect of undermining the usual proof of Cantor's paradox. If we could prove $A \prec \mathscr{P}(A)$, then, since $\mathscr{P}(V) = V$, we could obtain a contradiction from $V \prec \mathscr{P}(V)$. In NF, the standard proof of Cantor's theorem does yield $\mathrm{USC}(A) \prec \mathscr{P}(A)$, where $\mathrm{USC}(A)$ stands for $\{x \mid (\exists u)(u \in A \wedge x = \{u\})\}$. If we let $A = V$, we conclude that $\mathrm{USC}(V) \prec V$. Thus, V has the peculiar property that it is not equinumerous with the set of all unit sets of its elements. In NBG, the function f, defined by $f(u) = \{u\}$ for all u in A, establishes a one–one correspondence between A and $\mathrm{USC}(A)$ for any set A. However, the defining condition for f is not stratified, so that f may not exist in NF. If f does exist, A is said to be *strongly Cantorian*.

Other surprising properties of NF are the following.

1. The axiom of choice is disprovable in NF (Specker, 1953).
2. Any model for NF must be *non-standard* in the sense that a well-ordering of the finite cardinals or of the ordinals of the model is not possible in the metalanguage (Rosser and Wang, 1950).
3. The axiom of infinity is provable in NF (Specker, 1953).

Although property 3 would ordinarily be thought of as a great advantage, the fact of the provability of an axiom of infinity appeared to many logicians to be too strong a result. If *that* can be proved, then probably anything can be proved, that is, NF is likely to be inconsistent. In addition, the disprovability of the axiom of choice seems to make NF a poor choice for practising mathematicians. However, if we restrict attention to so-called *Cantorian* sets, sets A for which A and $\mathrm{USC}(A)$ are equinumerous, then it might be consistent to assume the axiom of choice for Cantorian sets and to do mathematics within the universe of Cantorian sets.

NF has another attractive feature. A substantial part of category theory (see MacLane, 1971) can be developed in a straightforward way in NF, whereas this is not possible in ZF, NBG or MK. Since category theory has become an important branch of mathematics, this is a distinct advantage for NF.

If the system obtained from NF by assuming the existence of an inaccessible ordinal is consistent, then ZF is consistent (see Orey, 1956a; Collins 1955). If we add to NF the assumption of the existence of an infinite strongly Cantorian set, then Zermelo's set theory Z is consistent (see Rosser, 1954).

The question of whether the consistency of ZF implies the consistency of NF is still open (as is the question of the reverse implication).

Let ST^- be the simple theory of types ST without the axiom of infinity. Given any closed wf $\mathscr{B}$ of ST, let $\mathscr{B}^+$ denote the result of adding 1 to the types of all variables in $\mathscr{B}$. Let SP denote the theory obtained from ST^- by adding as axioms the wfs $\mathscr{B} \Leftrightarrow \mathscr{B}^+$ for all closed wfs $\mathscr{B}$. Specker (1958; 1962) proved that NF is consistent if and only if SP is consistent.

Let NFU denote the theory obtained from NF by restricting the extensionality axiom to non-empty sets:

$$\text{NF1}^* \qquad (\exists u)(u \in x) \wedge (\forall z)(z \in x \Leftrightarrow z \in y) \Rightarrow x = y$$

Jensen (1968–69) proved that NFU is consistent if and only if ST^- is consistent, and the equiconsistency continues to hold when both theories are supplemented by the axiom of infinity or by axioms of infinity and choice.

Discussions of NF may be found in Hatcher (1982) and Quine (1963). Forster (1983) gives a survey of more recent results.

Quine also proposed a system ML that is formally related to NF in much the same way that MK is related to ZF. The variables are capital italic letters $X, Y, Z, \ldots$; these variables are called *class variables*. We define $M(X)$, X is a *set*,† by $(\exists Y)(X \in Y)$, and we introduce lower-case italic letters $x, y, z, \ldots$ as variables restricted to sets. Equality is defined as in NBG: $X = Y$ for $(\forall Z)(Z \in X \Leftrightarrow Z \in Y)$. Then we introduce an axiom of equality:

$$\text{ML1:} \quad X = Y \wedge X \in Z \Rightarrow Y \in Z$$

There is an unrestricted comprehension axiom scheme:

$$\text{ML2:} \quad (\exists Y)(\forall x)(x \in Y \Leftrightarrow \mathscr{B}(x))$$

where $\mathscr{B}(x)$ is any wf of ML. Finally, we wish to introduce an axiom that has the same effect as the comprehension axiom scheme NF2:

$$\text{ML3:} \quad (\forall y_1) \ldots (\forall y_n)(\exists z)(\forall x)(x \in z \Leftrightarrow \mathscr{B}(x))$$

where $\mathscr{B}(x)$ is any stratified wf whose free variables are $x, y_1, \ldots, y_n (n \geq 0)$ and whose quantifiers are set quantifiers.

All theorems of NF are provable in ML. Hence, if ML is consistent, so is NF. The converse has been proved by Wang (1950). In fact, any closed wf of NF provable in ML is already provable in NF.

ML has the same advantages over NF that MK and NBG have over ZF: a greater ease and power of expression. Moreover, the natural numbers of ML behave much better than those of NF; the principle of mathematical induction can be proved in full generality in ML.

†Quine uses the word 'element' instead of 'set'.

The prime source for ML is Quine (1951).[†] Consult also Quine (1963) and Fraenkel, Bar-Hillel and Lévy (1973).

Set theory with urelements

The theories NBG, MK, ZF, NF and ML do not allow for objects that are not sets or classes. This is all well and good for mathematicians, since only sets or classes seem to be needed for dealing with mathematical concepts and problems. However, if set theory is to be a part of a more inclusive theory having to do with the natural or social sciences, we must permit reference to things like electrons, molecules, people, companies, etc., and to sets and classes that contain such things. Things that are not sets or classes are sometimes called *urelements*.[‡] We shall sketch a theory UR similar to NBG that allows for the existence of urelements.[§] Like NBG, UR will have a finite number of axioms.

The variables of UR will be the lower-case Latin boldface letters $\mathbf{x}_1, \mathbf{x}_2, \ldots$. (As usual, let us use $\mathbf{x}, \mathbf{y}, \mathbf{z}, \ldots$ to refer to arbitrary variables.) In addition to the binary predicate letter A_2^2 there will be a monadic predicate letter A_1^1. We abbreviate $A_2^2(\mathbf{x}, \mathbf{y})$ by $\mathbf{x} \in \mathbf{y}$, $\neg A_2^2(\mathbf{x}, \mathbf{y})$ by $\mathbf{x} \notin \mathbf{y}$, and $A_1^1(\mathbf{x})$ by Cls $(\mathbf{x})$. (Read 'Cls(**x**)' as '**x** is a class'.) To bring our notation into line with that of NBG, we shall use capital Latin letters as restricted variables for classes. Thus, $(\forall X)\mathscr{B}(X)$ stands for $(\forall \mathbf{x})$ (Cls $(\mathbf{x}) \Rightarrow \mathscr{B}(\mathbf{x})$), and $(\exists X)\mathscr{B}(X)$ stands for $(\exists \mathbf{x})$ (Cls(**x**) $\wedge\ \mathscr{B}(\mathbf{x})$). Let M(**x**) stand for Cls(**x**) $\wedge$ $(\exists \mathbf{y}(\mathbf{x} \in \mathbf{y})$, and read 'M(**x**)' as '**x** is a set'. As in NBG, use lower-case Latin letters as restricted variables for sets. Thus, $(\forall x)\mathscr{B}(x)$ stands for $(\forall \mathbf{x})$ (M(**x**) $\Rightarrow \mathscr{B}(\mathbf{x})$), and $(\exists x)\mathscr{B}(x)$ stands for $(\exists \mathbf{x})$ (M(**x**) $\wedge \mathscr{B}(\mathbf{x})$). Let Pr(**x**) stand for Cls(**x**) $\wedge$ $\neg$M(**x**), and read 'Pr(**x**)' as '**x** is a proper class'. Introduce Ur(**x**) as an abbreviation for $\neg$Cls(**x**), and read 'Ur(**x**)' as '**x** is an urelement'. Thus, the domain of any model for UR will be divided into two disjoint parts consisting of the classes and the urelements, and the classes are divided into sets and proper classes. Let El(**x**) stand for M(**x**) $\vee$ Ur(**x**), and read 'El(**x**)' as **x** is an element'. In our intended interpretation, sets and urelements are the objects that are elements (i.e., members) of classes.

[†]Quine's earlier version of ML, published in 1940, was proved inconsistent by Rosser (1942). The present version is due to Wang (1950).

[‡]'Ur' is a German prefix meaning *primitive, original* or *earliest*. The words 'individual' and 'atom' are sometimes used as synonyms for 'urelement'.

[§]Zermelo's 1908 axiomatization permitted urelements. Fraenkel was among the first to draw attention to the fact that urelements are not necessary for mathematical purposes (see Fraenkel, 1928, pp. 355f). Von Neumann's (1925; 1928) axiom systems excluded urelements.

Exercise

4.98 Prove: $\vdash_{UR} (\forall \mathbf{x})(\mathrm{El}(\mathbf{x}) \Leftrightarrow \neg \mathrm{Pr}(\mathbf{x}))$.

We shall define equality in a different way for classes and urelements.

DEFINITION $\mathbf{x} = \mathbf{y}$ is an abbreviation for:

$$[\mathrm{Cls}(\mathbf{x}) \wedge \mathrm{Cls}(\mathbf{y}) \wedge (\forall \mathbf{z})(\mathbf{z} \in \mathbf{x} \Leftrightarrow \mathbf{z} \in \mathbf{y})] \vee [\mathrm{Ur}(\mathbf{x}) \wedge \mathrm{Ur}(\mathbf{y}) \wedge (\forall \mathbf{z})(\mathbf{x} \in \mathbf{z} \Leftrightarrow \mathbf{y} \in \mathbf{z})]$$

Exercise

4.99 Prove :$\vdash_{UR} (\forall \mathbf{x})(\mathbf{x} = \mathbf{x})$.

AXIOM UR1

$$(\forall \mathbf{x})(\mathrm{Ur}(\mathbf{x}) \Rightarrow (\forall \mathbf{y})(\mathbf{y} \notin \mathbf{x})]$$

Thus, urelements have no members.

Exercise

4.100 Prove: $\vdash_{UR} (\forall \mathbf{x})(\forall \mathbf{y})(\mathbf{x} \in \mathbf{y} \Rightarrow \mathrm{Cls}(\mathbf{y}) \wedge \mathrm{El}(\mathbf{x}))$.

AXIOM UR2

$$(\forall X)(\forall Y)(\forall Z)(X = Y \wedge X \in Z \Rightarrow Y \in Z)$$

Exercise

4.101 Show:

(a) $\vdash_{UR} (\forall \mathbf{x})(\forall \mathbf{y})(\mathbf{x} = \mathbf{y} \Rightarrow (\forall \mathbf{z})(\mathbf{z} \in \mathbf{x} \Leftrightarrow \mathbf{z} \in \mathbf{y}))$
(b) $\vdash_{UR} (\forall \mathbf{x})(\forall \mathbf{y})(\mathbf{x} = \mathbf{y} \Rightarrow (\forall \mathbf{z})(\mathbf{x} \in \mathbf{z} \Leftrightarrow \mathbf{y} \in \mathbf{z}))$
(c) $\vdash_{UR} (\forall \mathbf{x})(\forall \mathbf{y})(\mathbf{x} = \mathbf{y} \Rightarrow [\mathrm{Cls}(\mathbf{x}) \Leftrightarrow \mathrm{Cls}(\mathbf{y})] \wedge [\mathrm{Ur}(\mathbf{x}) \Leftrightarrow \mathrm{Ur}(\mathbf{y})] \wedge \mathrm{M}(\mathbf{x}) \Leftrightarrow \mathrm{M}(\mathbf{y})])$
(d) $\vdash_{UR} (\forall \mathbf{x})(\forall \mathbf{y})[\mathbf{x} = \mathbf{y} \Rightarrow (\mathscr{B}(\mathbf{x}, \mathbf{x}) \Rightarrow \mathscr{B}(\mathbf{x}, \mathbf{y}))]$, where $\mathscr{B}(\mathbf{x}, \mathbf{y})$ arises from $\mathscr{B}(\mathbf{x}, \mathbf{x})$ by replacing some, but not necessarily all, free occurrences of $\mathbf{x}$ by $\mathbf{y}$, with the proviso that $\mathbf{y}$ is free for $\mathbf{x}$ in $\mathscr{B}(\mathbf{x}, \mathbf{x})$.
(e) UR is a first-order theory with equality (with respect to the given definition of equality).

AXIOM UR3 (NULL SET)

$$(\exists x)(\forall \mathbf{y})(\mathbf{y} \notin x)$$

This tell us that there is a *set* that has no members. Of course, all urelements also have no elements.

Exercise

4.102 Show:
$\vdash_{UR} (\exists_1 x)(\forall \mathbf{y})(\mathbf{y} \notin x)$. On the basis of this exercise we can introduce a new individual constant $\emptyset$ satisfying the condition $\mathrm{M}(\emptyset) \wedge (\forall \mathbf{y})(\mathbf{y} \notin \emptyset)$.

AXIOM UR4 (PAIRING)

$$(\forall \mathbf{x})(\forall \mathbf{y})(\mathrm{El}(\mathbf{x}) \wedge \mathrm{El}(\mathbf{y}) \Rightarrow (\exists z)(\forall \mathbf{u})(\mathbf{u} \in z \Leftrightarrow [\mathbf{u} = \mathbf{x} \vee \mathbf{u} = \mathbf{y}])$$

Exercise

4.112 Prove: $\vdash_{UR} (\forall \mathbf{x})(\forall \mathbf{y})(\exists_1 z)([\mathrm{El}(\mathbf{x}) \wedge \mathrm{El}(\mathbf{y}) \wedge (\forall \mathbf{u})(\mathbf{u} \in z \Leftrightarrow [\mathbf{u} = \mathbf{x} \vee \mathbf{u} = \mathbf{y}]) \vee [(\neg \mathrm{El}(\mathbf{x}) \vee \neg \mathrm{El}(\mathbf{y})) \wedge z = \emptyset])$

On the basis of this exercise we can introduce the unordered pair notation $\{\mathbf{x}, \mathbf{y}\}$. When $\mathbf{x}$ and $\mathbf{y}$ are elements, $\{\mathbf{x}, \mathbf{y}\}$ is the set that has $\mathbf{x}$ and $\mathbf{y}$ as its only members; when $\mathbf{x}$ or $\mathbf{y}$ is a proper class, $\{\mathbf{x}, \mathbf{y}\}$ is arbitrarily chosen to be the empty set $\emptyset$. As usual, the singleton notation $\{\mathbf{x}\}$ stands for $\{\mathbf{x}, \mathbf{x}\}$.

DEFINITION (ORDERED PAIR)

Let $\langle \mathbf{x}, \mathbf{y} \rangle$ stand for $\{\{\mathbf{x}\}, \{\mathbf{x}, \mathbf{y}\}\}$. As in the proof of Proposition 4.3, one can show that, for any elements $\mathbf{x}, \mathbf{y}, \mathbf{u}, \mathbf{v}$, $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle \Leftrightarrow [\mathbf{x} = \mathbf{u} \wedge \mathbf{y} = \mathbf{v}]$. Ordered n-tuples can be defined as in NBG.

The class existence axioms B1–B7 of NBG have to be altered slightly by sometimes replacing universal quantification with respect to sets by universal quantification with respect to elements.

AXIOMS OF CLASS EXISTENCE

(UR5) $(\exists X)(\forall \mathbf{u})(\forall \mathbf{v})(\mathrm{El}(\mathbf{u}) \wedge \mathrm{El}(\mathbf{v}) \Rightarrow [\langle \mathbf{u}, \mathbf{v} \rangle \in X \Leftrightarrow \mathbf{u} \in \mathbf{v}])$
(UR6) $(\forall X)(\forall Y)(\exists Z)(\forall \mathbf{u})(\mathbf{u} \in Z \Leftrightarrow \mathbf{u} \in X \wedge \mathbf{u} \in Y)$
(UR7) $(\forall X)(\exists Z)(\forall \mathbf{u})(\mathrm{El}(\mathbf{u}) \Rightarrow [\mathbf{u} \in Z \Leftrightarrow \mathbf{u} \notin X])$
(UR8) $(\forall X)(\exists Z)(\forall \mathbf{u})(\mathrm{El}(\mathbf{u}) \Rightarrow (\mathbf{u} \in Z \Leftrightarrow (\exists \mathbf{v})(\langle \mathbf{u}, \mathbf{v} \rangle \in X))$

(UR9) $(\forall X)(\exists Z)(\forall \mathbf{u})(\forall \mathbf{v})(\mathrm{El}(\mathbf{u}) \wedge \mathrm{El}(\mathbf{v}) \Rightarrow (\langle \mathbf{u}, \mathbf{v} \rangle \in Z \Leftrightarrow \mathbf{u} \in X))$

(UR10) $(\forall X)(\exists Z)(\forall \mathbf{u})(\forall \mathbf{v})(\forall \mathbf{w})(\mathrm{El}(\mathbf{u}) \wedge \mathrm{El}(\mathbf{v}) \wedge \mathrm{El}(\mathbf{w}) \Rightarrow [\langle \mathbf{u}, \mathbf{v}, \mathbf{w} \rangle \in Z \Leftrightarrow \langle \mathbf{v}, \mathbf{w}, \mathbf{u} \rangle \in X])$

(UR11) $(\forall X)(\exists Z)(\forall \mathbf{u})(\forall \mathbf{v})(\forall \mathbf{w})(\mathrm{El}(\mathbf{u}) \wedge \mathrm{El}(\mathbf{v}) \wedge \mathrm{El}(\mathbf{w}) \Rightarrow [\langle \mathbf{u}, \mathbf{v}, \mathbf{w} \rangle \in Z \Leftrightarrow \langle \mathbf{u}, \mathbf{w}, \mathbf{v} \rangle \in X])$

As in NBG, we can prove the existence of the intersection, complement and union of any classes, and the existence of the class V of all elements. But in UR we also need an axiom to ensure the existence of the class V_{M} of all sets, or, equivalently, of the class V_{ur} of all urelements.

AXIOM UR12

$$(\exists X)(\forall \mathbf{u})(\mathbf{u} \in X \Leftrightarrow \mathrm{Ur}(\mathbf{u}))$$

This yields the existence of V_{ur} and implies the existence of V_{M}, that is, $(\exists X)(\forall \mathbf{u})(\mathbf{u} \in X \Leftrightarrow \mathrm{M}(\mathbf{u}))$. The class V_{El} of all elements is then the union $V_{\mathrm{ur}} \cup V_{\mathrm{M}}$. Note that this axiom also yields $(\exists X)(\forall \mathbf{u})(\mathrm{El}(\mathbf{u}) \Rightarrow [\mathbf{u} \in X \Leftrightarrow \mathrm{Cls}(\mathbf{u})])$, since V_{M} can be taken as the required class X.

As in NBG, we can prove a general class existence theorem.

Exercise

4.104. Let $\varphi(\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{y}_1, \ldots, \mathbf{y}_m)$ be a formula in which quantification takes place only with respect to elements, that is, any subformula $(\forall \mathbf{u})\mathscr{B}$ has the form $(\forall \mathbf{u})(\mathrm{El}(\mathbf{u}) \Rightarrow \mathscr{C})$. Then

$$\vdash_{\mathrm{UR}} (\exists Z)(\forall \mathbf{x}_1) \ldots (\forall \mathbf{x}_n)(\mathrm{El}(\mathbf{x}_1) \wedge \ldots \wedge \mathrm{El}(\mathbf{x}_n) \Rightarrow [\langle \mathbf{x}_1, \ldots, \mathbf{x}_n \rangle \in Z \Leftrightarrow \varphi(\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{y}_1, \ldots, \mathbf{y}_m)]).$$

The sum set, power set, replacement and infinity axioms can be translated into UR.

AXIOM UR13

$$(\forall x)(\exists y)(\forall \mathbf{u})(\mathbf{u} \in y \Leftrightarrow (\exists \mathbf{v})(\mathbf{u} \in \mathbf{v} \wedge \mathbf{v} \in x))$$

AXIOM UR14

$$(\forall x)(\exists y)(\forall \mathbf{u})(\mathbf{u} \in y \Leftrightarrow \mathbf{u} \subseteq x)$$

where $\mathbf{u} \subseteq \mathbf{x}$ stands for $\mathrm{M}(\mathbf{u}) \wedge \mathrm{M}(\mathbf{x}) \wedge (\forall \mathbf{v})(\mathbf{v} \in \mathbf{u} \Rightarrow \mathbf{v} \in \mathbf{x})$.

AXIOM UR15

$$(\forall Y)(\forall x)(\mathrm{Un}(Y) \Rightarrow (\exists y)(\forall \mathbf{u})[\mathbf{u} \in y \Leftrightarrow (\exists \mathbf{v})(\langle \mathbf{v}, \mathbf{u} \rangle \in Y \wedge \mathbf{v} \in x)])$$

where $\mathrm{Un}(\mathbf{z})$ stands for $(\forall \mathbf{x}_1)(\forall \mathbf{x}_2)(\forall \mathbf{x}_3)[\mathrm{El}(\mathbf{x}_1) \wedge \mathrm{El}(\mathbf{x}_2) \wedge \mathrm{El}(\mathbf{x}_3) \Rightarrow (\langle \mathbf{x}_1, \mathbf{x}_2 \rangle \in \mathbf{z} \wedge \langle \mathbf{x}_1, \mathbf{x}_3 \rangle \mathbf{z} \Rightarrow \mathbf{x}_2 = \mathbf{x}_3)]$

AXIOM UR16

$$(\exists x)(\emptyset \in x \wedge (\forall u)(u \in x \Rightarrow u \cup \{u\} \in x))$$

From this point on, the standard development of set theory including the theory of ordinal numbers, can be imitated in UR.

PROPOSITION 4.49

NBG is a subtheory of UR.

Proof

It is easy to verify that every axiom of NBG is provable in UR, provided that we take the variables of NBG as restricted variables for 'classes' in UR. The restricted variables for sets in NBG become restricted variables for 'sets' in UR.†

PROPOSITION 4.50

UR is consistent if and only if NBG is consistent.

Proof

By Proposition 4.49, if UR is consistent, NBG is consistent. For the converse, note that any model of NBG yields a model of UR in which there are no urelements. In fact, if we replace '$\mathrm{Cls}(\mathbf{x})$' by the NBG formula '$x = x$', then the axioms of UR become theorems of NBG. Hence, a proof of a contradiction in UR would produce a proof of a contradiction in NBG.

The axiom of regularity (Reg) takes the following form in UR.

†In fact, a formula $(\forall x)\mathscr{B}(x)$ in NBG is an abbreviation in NBG for $(\forall X)((\exists Y)(X \in Y) \Rightarrow \mathscr{B}(X))$. The latter formula is equivalent in UR to $(\forall \mathbf{x})(\mathrm{M}(\mathbf{x}) \Rightarrow \mathscr{B}(\mathbf{x}))$, which is abbreviated as $(\forall x)\mathscr{B}(x)$ in UR.

$$(\mathbf{Reg_{UR}})\ (\forall X)(X \neq \emptyset \Rightarrow (\exists \mathbf{u})(\mathbf{u} \in X \wedge \neg(\exists \mathbf{v})(\mathbf{v} \in X \wedge \mathbf{v} \in \mathbf{u})))$$

It is clear that an analogue of Proposition 4.49 holds: UR + $(\mathbf{Reg_{UR}})$ is an extension of NBG + (Reg). Likewise, the argument of Proposition 4.50 shows the equiconsistency of NBG + (Reg) and UR + $(\mathbf{Reg_{UR}})$.

Since definition by transfinite induction (Proposition 4.14(b)) holds in UR, the cumulative hierarchy can be defined

$$\Psi`\emptyset = \emptyset$$
$$\Psi`(\alpha') = \mathscr{P}(\Psi`\alpha)$$
$$\mathrm{Lim}(\lambda) \Rightarrow \Psi`\lambda = \bigcup_{\beta <_o \lambda} \Psi`\beta$$

and the union $H = \bigcup(\Psi``On)$ is the class of 'pure' sets in UR and forms a model of NBG + (Reg). In NBG, by Proposition 4.45, (Reg) is equivalent to $V = H$, where V is the class of all sets.

If the class V_{ur} of urelements is a set, then we can define the following by transfinite induction:

$$\Xi`\emptyset = V_{\mathrm{ur}}$$
$$\Xi`(\alpha') = \mathscr{P}(\Xi`\alpha)$$
$$\mathrm{Lim}(\lambda) \Rightarrow \Xi`\lambda = \bigcup_{\beta <_o \lambda} \Xi`\beta$$

The union $H_{\mathrm{ur}} = \bigcup(\Xi``On)$ is a model of UR + $(\mathbf{Reg_{UR}})$, and $(\mathbf{Reg_{UR}})$ holds in UR if and only if H_{ur} is the class V_{El} of all elements.

If the class V_{ur} of urelements is a proper class, it is possible to obtain an analogue of H_{ur} in the following way. For any set $\mathbf{x}$ whose members are urelements and any ordinal γ, we can define a function $\Xi^{\gamma}_{\mathbf{x}}$ by transfinite induction up to γ:

$$\Xi^{\gamma}_{\mathbf{x}}`\emptyset = \mathbf{x}$$
$$\Xi^{\gamma}_{\mathbf{x}}`(\alpha') = \mathscr{P}(\Xi^{\gamma}_{\mathbf{x}}`\alpha) \qquad \text{if } \alpha' <_o \gamma$$
$$\mathrm{Lim}(\lambda) \Rightarrow \Xi^{\gamma}_{\mathbf{x}}`\lambda = \bigcup_{\beta <_o \lambda} \Xi^{\gamma}_{\mathbf{x}}`\beta \qquad \text{if } \lambda <_o \gamma$$

Let H^*_{ur} be the class of all elements $\mathbf{v}$ such that, for some $\mathbf{x}$ and γ, $\mathbf{v}$ is in the range of $\Xi^{\gamma}_{\mathbf{x}}$. Then H^*_{ur} determines a model of UR + $(\mathbf{Reg_{UR}})$, and, in UR, $(\mathbf{Reg_{UR}})$ holds if and only if H^*_{ur} is the class V_{El} of all elements.

The equiconsistency of NBG and UR can be strengthened to show the following result.

PROPOSITION 4.51

If NBG is consistent, then so is the theory UR + $(\mathbf{Reg_{ur}})$ + 'V_{ur} is denumerable'.

Proof

Within NBG one can define a model with domain ω that is a model of NBG without the axiom of infinity. The idea is due to Ackermann (1937). For any n and m in ω, define $m \mathrel{\tilde{\in}} n$ to mean that 2^m occurs as a term in the expansion of n as a sum of different powers of 2.† If we take 'A-sets' to be members of ω and 'proper A-classes' to be infinite subsets of ω, it is easy to verify all axioms of NBG + (Reg) except the axiom of infinity.‡ (See Bernays (1954, pp. 81–82) for a sketch of the argument.) Then we change the 'membership' relation on ω by defining $m \in_1 n$ to mean that $2^m \tilde{\in} n$. Now we define a 'set' to be either 0 or a member n of ω for which there is some m in ω such that $m \in_1 n$. We take the 'urelements' to be the members of ω that are not 'sets'. For example, 8 is an 'urelement', since $8 = 2^3$ and 3 is not a power of 2. Other small 'urelements' are 1, 9, 32, 33 and 40. In general, the 'urelements' are sums of one or more distinct powers 2^k, where k is not a power of 2. The 'proper classes' are to be the infinite subsets of ω. Essentially the same argument as for Ackermann's model shows that this yields a model $\mathscr{M}$ of all axioms of UR + ($\mathbf{Reg_{UR}}$) except the axiom of infinity. Now we want to extend $\mathscr{M}$ to a model of UR. First, let r stand for the set of all finite subsets of ω that are not members of ω, and then define by transfinite induction the following function Θ.

$$\Theta`\emptyset = \omega$$
$$\Theta`(\alpha') = \mathscr{P}(\Theta`\alpha) - r$$
$$\mathrm{Lim}(\lambda) \Rightarrow \Theta`\lambda = \bigcup_{\beta <_o \lambda} \Theta`\beta$$

Let $H_B = \bigcup(\Theta``On)$. Note that H_B contains no members of r. Let us define a membership relation $\in^*$ on H_B. For any members x and y of H_B, define $x \in^* y$ to mean that either x and y are in ω and $x \in_1 y$, or $y \notin \omega$ and $x \in y$. The 'urelements' will be those members of ω that are the 'urelements' of $\mathscr{M}$. The 'sets' will be the ordinary sets of H_B that are not 'urelements', and the 'proper classes' will be the proper classes of NBG that are subclasses of H_B. It now requires a long careful argument to show that we have a model of UR + ($\mathbf{Reg_{UR}}$) in which the class of urelements is a denumerable set.

A uniform method for constructing a model of UR + ($\mathbf{Reg_{UR}}$) in which the class of urelements is a set of arbitrary size may be found in Brunner (1990, p.65).§ If AC holds in the underlying theory, it holds in the model as well.

†This is equivalent to the statement that the greatest integer k such that $k \cdot 2^m \leqslant n$ is odd.

‡For distinct natural numbers $n_1, \ldots, n_k$, the role of the finite set $\{n_1, \ldots, n_k\}$ is played by the natural number $2^{n_1} + \ldots + 2^{n_k}$.

§Brunner attributes the idea behind the construction to J. Truss.

The most important application of axiomatic set theories with urelements used to be the construction of independence proofs. The first independence proof for the axiom of choice, given by Fraenkel (1922b), depended essentially on the existence of a denumerable set of urelements. More precise formulations and further developments may be found in Lindenbaum and Mostowski (1938) and Mostowski (1939).† Translations of these proofs into set theories without urelements were found by Shoenfield (1955), Mendelson (1956b) and Specker (1957), but only at the expense of weakening the axiom of regularity. This shortcoming was overcome by the forcing method of Cohen (1966), which applies to theories with (Reg) and without urelements.

†For more information about these methods, see Levy (1965), Pincus (1972), Howard (1973) and Brunner (1990).

Computability

5

5.1 ALGORITHMS. TURING MACHINES

An *algorithm* is a computational method for solving each and every problem from a large class of problems. The computation has to be precisely specified so that it requires no ingenuity for its performance. The familiar technique for adding integers is an algorithm, as are the techniques for computing the other arithmetic operations of subtraction, multiplication and division. The truth table procedure to determine whether a statement form is a tautology is an algorithm within logic itself.

It is often easy to see that a specified procedure yields a desired algorithm. In recent years, however, many classes of problems have been proved not to have an algorithmic solution. Examples are:

1. Is a given wf of quantification theory logically valid?
2. Is a given wf of formal number theory S true (in the standard interpretation)?
3. Is a given wf of S provable in S?
4. Does a given polynomial $f(x_1, \ldots, x_n)$ with integral coefficients have integral roots (Hilbert's tenth problem)?

In order to prove rigorously that there does *not* exist an algorithm for answering such questions, it is necessary to supply a precise definition of the notion of algorithm.

Various proposals for such a definition were independently offered in 1936 by Church (1936b), Turing (1936–37), and Post (1936). All of these definitions, as well as others proposed later, have been shown to be equivalent. Moreover, it is intuitively clear that every procedure given by these definitions is an algorithm. On the other hand, every known algorithm falls under these definitions. Our exposition will use Turing's ideas.

First of all, the objects with which an algorithm deals may be assumed to be the symbols of a finite alphabet $A = \{a_0, a_1, \ldots, a_n\}$. Non-symbolic

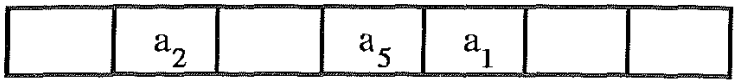


Figure 5.1

objects can be represented by symbols, and languages actually used for computation require only finitely many symbols.†

A finite sequence of symbols of a language A is called a *word* of A. It is convenient to admit an empty word Λ consisting of no symbols at all. If P and Q are words, then PQ denotes the word obtained by writing Q to the right of P. For any positive integer k, P^k shall stand for the word made up of k consecutive occurrences of P.

The work space of an algorithm often consists of a piece of paper or a blackboard. However, we shall make the simplifying assumption that all calculations take place on a tape that is divided into squares (see Figure 5.1). The tape is potentially infinite in both directions in the sense that, although at any moment it is finite, more squares always can be added to the right- and left-hand ends of the tape. Each square contains at most one symbol of the alphabet A. At any one time, only a finite number of squares contain symbols, while the rest are blank. The symbol a_0 will be reserved for the content of a blank square. (In ordinary language, a space is sometimes used for the same purpose.) Thus, the condition of the tape at a given moment can be represented by a word of A; the tape in Figure 5.1 is $a_2a_0a_5a_1$. Our use of a one-dimensional tape does not limit the algorithms that can be handled; the information in a two-dimensional array can be encoded as a finite sequence.‡

Our computing device, which we shall refer to as a *Turing machine*, works in the following way. The machine operates at discrete moments of time, not continuously. It has a *reading head* which, at any moment, will be scanning one square of the tape. (Observation of a larger domain could be reduced to consecutive observations of individual squares.) The device then reacts in any of four different ways:

1. It prints a symbol in the square, erasing the previous symbol.
2. It moves to the next square to the right.
3. It moves to the next square to the left.
4. It stops.

†If a language has a denumerable alphabet $\{a_0, a_1, \ldots\}$, then we can replace it by the alphabet $\{b, *\}$. Each symbol a_n of the old alphabet can be replaced by the expression $b*\cdots*$, consisting of b followed by n occurrences of $*$.

‡This follows from the fact that there is an effective one–one correspondence between the set of pairs of natural numbers and the set of natural numbers. For the details, see pp. 183–4.

What the machine does depends not only on the observed symbol but also on the *internal state* of the machine at that moment (which, in turn, depends on the previous steps of the computation and on the structure of the machine). We shall make the plausible assumption that a machine has only a finite number of internal states $\{q_0, q_1, \ldots, q_m\}$. The machine will always begin its operation in the *initial state* q_0.

A step in a computation corresponds to a quadruple of one of the following three forms: (1) $q_j a_i a_k q_r$; (2) $q_j a_i R q_r$; (3) $q_j a_i L q_r$. In each case, q_j is the present internal state, a_i is the symbol being observed, and q_r is the internal state after the step. In form (1), the machine erases a_i and prints a_k. In form (2), the reading head of the machine moves one square to the right, and, in form (3), it moves one square to the left. We shall indicate later how the machine is told to stop.

Now we can give a precise definition. A *Turing machine* with an alphabet A of *tape symbols* $\{a_0, a_1, \ldots, a_n\}$ and with *internal states* $\{q_0, q_1, \ldots, q_m\}$ is a finite set $\mathcal{T}$ of quadruples of the forms (1) $q_j a_i a_k q_r$, (2) $q_j a_i R q_r$, and (3) $q_j a_i L q_r$ such that no two quadruples of $\mathcal{T}$ have the same first two symbols.

Thus, for fixed $q_j a_i$, no two quadruples of types (1), (2) and (3) are in $\mathcal{T}$. This condition ensures that there is never a situation in which the machine is instructed to perform two contradictory operations.

The Turing machine $\mathcal{T}$ operates in accordance with its list of quadruples. This can be made precise in the following manner.

By a *tape description* of $\mathcal{T}$ we mean a word such that: (1) all symbols in the word but one are tape symbols; (2) the only symbol that is not a tape symbol is an internal state q_j; and (3) q_j is not the last symbol of the word.

A tape description describes the condition of the machine and the tape at a given moment. When read from left to right, the tape symbols in the description represent the symbols on the tape at that moment, and the tape symbol that occurs immediately to the right of q_j in the tape description represents the symbol being scanned by the reading head at that moment. If the internal state q_j is the initial state q_0, then the tape description is called an *initial* tape description.

Example

The tape description $a_2 a_0 q_1 a_0 a_1 a_1$ indicates that the machine is in the internal state q_1, the tape is as shown in Figure 5.2, and the reading head is scanning the square indicated by the arrow.

We say that $\mathcal{T}$ *moves* one tape description α into another one β (abbreviated $\alpha \underset{\mathcal{T}}{\twoheadrightarrow} \beta$) if and only if one of the following is true.

1. α is of the form $Pq_j a_i Q$, β is of the form $Pq_r a_k Q$, and $q_j a_i a_k q_r$ is one of the quadruples of $\mathcal{T}$.†

†Here and below, P and Q are arbitrary (possibly empty) words of the alphabet of $\mathcal{T}$.

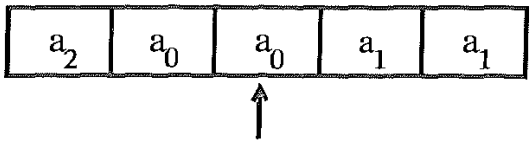


Figure 5.2

2. α is of the form $Pa_sq_ja_iQ$, β is of the form $Pq_ra_sa_iQ$, and $q_ja_iLq_r$ is one of the quadruples of $\mathscr{T}$.
3. α is of the form q_ja_iQ, β is of the form $q_ra_0a_iQ$, and $q_ja_iLq_r$ is one of the quadruples of $\mathscr{T}$.
4. α is of the form $Pq_ja_ia_kQ$, β is of the form $Pa_iq_ra_kQ$, and $q_ja_iRq_r$ is one of the quadruples of $\mathscr{T}$.
5. α is of the form Pq_ja_i, β is of the form $Pa_iq_ra_0$, and $q_ja_iRq_r$ is one of the quadruples of $\mathscr{T}$.

According to our intuitive picture, '$\mathscr{T}$ moves α into β' means that, if the condition at a time t of the Turing machine and tape is described by α, then the condition at time $t+1$ is described by β. Notice that, by clause 3, whenever the machine reaches the left-hand end of the tape and is ordered to move left, a blank square is attached to the tape on the left; similarly, by clause 5, a blank square is added on the right when the machine reaches the right-hand end and has to move right.

We say that $\mathscr{T}$ *stops* at tape description α if and only if there is no tape description β such that $\alpha \underset{\mathscr{T}}{\rightarrow\!\!\!\!\rightarrow} \beta$. This happens when q_ja_i occurs in α but q_ja_i is not the beginning of any quadruple of $\mathscr{T}$.

A *computation* of $\mathscr{T}$ is a finite sequence of tape descriptions $\alpha_0, \ldots, \alpha_k$ $(k \geq 0)$ such that the following conditions hold.

1. α_0 is an initial tape description, that is, the internal state occurring in α is q_0.
2. $\alpha_i \underset{\mathscr{T}}{\rightarrow\!\!\!\!\rightarrow} \alpha_{i+1}$ for $0 \leqslant i < k$
3. $\mathscr{T}$ stops at α_k.

This computation is said to *begin* at α_0 and *end* at α_k. If there is a computation beginning at α_0, we say that $\mathscr{T}$ is *applicable* to α_0.

The algorithm $\mathrm{Alg}_{\mathscr{T}}$ determined by $\mathscr{T}$ is defined as follows:

> For any words P and Q of the alphabet A of $\mathscr{T}$, $\mathrm{Alg}_{\mathscr{T}}(P) = Q$ if and only if there is a computation of $\mathscr{T}$ that begins with the tape description q_0P and ends with a tape description of the form $R_1q_jR_2$, where $Q = R_1R_2$.

This means that, when $\mathscr{T}$ begins at the left-hand end of P and there is nothing else on the tape, $\mathscr{T}$ stops with Q as the entire content of the tape. Notice that $\mathrm{Alg}_{\mathscr{T}}$ need not be defined for certain words P. An algorithm $\mathrm{Alg}_{\mathscr{T}}$ determined by a Turing machine $\mathscr{T}$ is said to be a Turing *algorithm*.

Example
In any computation of the Turing machine $\mathscr{T}$ given by

$$q_0a_0Rq_0, q_0a_1a_0q_1, q_0a_2a_0q_1, \ldots, q_0a_na_0q_1$$

$\mathscr{T}$ locates the first non-blank symbol (if any) at or to the right of the square scanned at the beginning of the computation, erases that symbol, and then stops. If there are only blank squares at or to the right of the initial square, $\mathscr{T}$ keeps on moving right for ever.

Let us now consider computations of number-theoretic functions. For convenience, we sometimes will write | instead of a_1 and B instead of a_0. (Think of B as standing for 'blank'.) For any natural number k, its *tape representation* $\overline{k}$ will stand for the word $|^{k+1}$, that is, the word consisting of $k+1$ occurrences of |. Thus, $\overline{0} = |, \overline{1} = ||, \overline{2} = |||$, and so on. The reason why we represent k by $k+1$ occurrences of | instead of k occurrences is that we wish $\overline{0}$ to be a non-empty word, so that we will be aware of its presence. The tape representation $\overline{(k_1, k_2, \ldots, k_n)}$ of an n-tuple of natural numbers $(k_1, k_2, \ldots, k_n)$ is defined to be the word $\overline{k_1}\mathrm{B}\ \overline{k_2}\mathrm{B}\cdots\mathrm{B}\overline{k_n}$. For example, $\overline{(3,1,0,5)}$ is ||||B||B|B||||||.

A Turing machine $\mathscr{T}$ will be thought of as computing the following partial function $f_{\mathscr{T},1}$ of one variable.†

> $f_{\mathscr{T},1}(k) = m$ if and only if the following condition holds: $\mathrm{Alg}_{\mathscr{T}}(\overline{k})$ is defined and $\mathrm{Alg}_{\mathscr{T}}(\overline{k}) = \mathrm{E}_1\,\overline{m}\,\mathrm{E}_2$, where E_1 and E_2 are certain (possibly empty) words consisting of only Bs (blanks).

The function $f_{\mathscr{T},1}$ is said to be *Turing-computable*. Thus, a one-place partial function f is Turing-computable if and only if there is a Turing machine such that $f = f_{\mathscr{T},1}$.

For each $n > 1$, a Turing machine $\mathscr{T}$ also computes a partial function $f_{\mathscr{T},n}$ of n variables. For any natural numbers $k_1, \ldots, k_n$:

> $f_{\mathscr{T},n}(k_1, \ldots, k_n) = m$ if and only if the following condition holds:
>
> $\mathrm{Alg}_{\mathscr{T}}(\overline{(k_1, k_2, \ldots, k_n)})$ is defined and $\mathrm{Alg}_{\mathscr{T}}(\overline{(k_1, k_2, \ldots, k_n)}) = \mathrm{E}_1\,\overline{m}\,\mathrm{E}_2$, where E_1 and E_2 are certain (possibly empty) words consisting of only Bs (blanks).

The partial function $f_{\mathscr{T},n}$ is said to be *Turing-computable*. Thus, an n-place partial function f is Turing-computable if and only if there is a Turing machine $\mathscr{T}$ such that $f = f_{\mathscr{T},n}$.

Notice that, at the end of a computation of a value of a Turing-computable function, only the value appears on the tape, aside from blank squares at either or both ends, and the location of the reading head does not matter. Also observe that, whenever the function is not defined, either the

†Remember that a partial function *may* fail to be defined for some values of its argument. Thus, a total function is considered to be a special case of a partial function.

Turing machine will never stop or, if it does stop, the resulting tape is not of the appropriate form $E_1 \overline{m} E_2$.

Examples

1. Consider the Turing machine $\mathcal{T}$, with alphabet $\{B, |\}$, defined by $q_0|Lq_1, q_1B|q_2$. $\mathcal{T}$ computes the successor function $N(x)$, since $q_0\overline{k} \underset{\mathcal{T}}{\twoheadrightarrow} q_1\ B\overline{k} \underset{\mathcal{T}}{\twoheadrightarrow} q_2\overline{k+1}$, and $\mathcal{T}$ stops at $q_2\overline{k+1}$. Hence $N(x)$ is Turing-computable.
2. The Turing machine $\mathcal{T}$ defined by

$$q_0|Bq_1, q_1BRq_0, q_0B|q_2$$

computes the zero function $Z(x)$. Given $\overline{k}$ on the tape, $\mathcal{T}$ moves right, erasing all |s until it reaches a blank, which it changes to a|. So, $\overline{0}$ is the final result. Thus, $Z(x)$ is Turing-computable.

3. The addition function is computed by the Turing machine $\mathcal{T}$ defined by the following seven quadruples:

$$q_0|Bq_0, q_0BRq_1, q_1|Rq_1, q_1B|q_2, q_2|Rq_2, q_2BL|q_3, q_3|Bq_3$$

In fact, for any natural numbers m and n,

$$q_0\overline{(m,n)} = q_0|^{m+1}B|^{n+1} \underset{\mathcal{T}}{\twoheadrightarrow} q_0B|^mB|^{n+1} \underset{\mathcal{T}}{\twoheadrightarrow} Bq_1|^mB|^{n+1}$$

$$\underset{\mathcal{T}}{\twoheadrightarrow} \cdots \underset{\mathcal{T}}{\twoheadrightarrow} B|^mq_1B|^{n+1} \underset{\mathcal{T}}{\twoheadrightarrow} B|^mq_2||^{n+1} \underset{\mathcal{T}}{\twoheadrightarrow} \cdots$$

$$\underset{\mathcal{T}}{\twoheadrightarrow} B|^m|^{n+2}q_2B \underset{\mathcal{T}}{\twoheadrightarrow} B|^{m+n+1}q_3|B \underset{\mathcal{T}}{\twoheadrightarrow} B|^{m+n+1}q_3BB = B\overline{m+n}q_3BB$$

and $\mathcal{T}$ stops at $B\overline{m+n}q_3BB$.

Exercises

5.1 Show that the function U_2^2 such that $U_2^2(x_1, x_2) = x_2$ is Turing-computable.

5.2 (a) What function $f(x_1, x_2, x_3)$ is computed by the following Turing machine?

$$q_0||q_1, q_1|Bq_0, q_0BRq_1, q_1BRq_2,$$
$$q_2|Rq_2, q_2BRq_3, q_3|Bq_4, q_4BRq_3$$

(b) What function $f(x)$ is computed by the following Turing machine?

$$q_0|Bq_1, q_1BRq_2, q_2B|q_2$$

5.3 (a) State in plain language the operation of the Turing machine, described in Example 3, for computing the addition function.

(b) Starting with the tape description $q_0|||B||||$, write the sequence of tape descriptions that make up the computation by the addition machine of Example 3.

5.4 What function $f(x)$ is computed by the following Turing machine?

$q_0|Rq_1$ $q_4|Rq_4$ $q_6B|q_0$
$q_1|Bq_2$ $q_4B|q_5$ $q_1B|q_7$
q_2BRq_3 $q_5|Lq_5$ $q_7|Lq_7$
$q_3|Rq_3$ q_5BLq_6 q_7BRq_8
q_3BRq_4 $q_6|Lq_6$ $q_8|Bq_8$

5.5 Find a Turing machine that computes the function $sg(x)$. (Recall that $sg(0) = 0$ and $sg(x) = 1$ for $x > 0$.)

5.6^D Find Turing machines that compute the following functions.

(a) $x \dot{-} y$ (Remember that $x \dot{-} y = x - y$ if $x \geq y$, and $x \dot{-} y = 0$ if $x < y$.)

(b) $[x/2]$ (Recall that $[x/2]$ is the greatest integer less than or equal to $x/2$. Thus, $[x/2] = x/2$ if x is even, and $[x/2] = (x-1)/2$ if x is odd.)

(c) $x \cdot y$, the product of x and y.

5.7 If a function is Turing-computable, show that it is computable by infinitely many different Turing machines.

5.2 DIAGRAMS

Many Turing machines that compute even relatively simple functions (like multiplication) require a large number of quadruples. It is difficult and tedious to construct such machines, and even more difficult to check that they do the desired job. We shall introduce a pictorial technique for constructing Turing machines so that their operation is easier to comprehend. The basic ideas and notation are due to Hermes (1965).

1. Let $\mathscr{T}_1, \ldots, \mathscr{T}_r$ be any Turing machines with alphabet $A = \{a_0, a_1, \ldots, a_k\}$.
2. Select a finite set of points in a plane. These points will be called *vertices*.
3. To each vertex attach the name of one of the machines $\mathscr{T}_1, \ldots, \mathscr{T}_r$. Copies of the same machine may be assigned to more than one vertex.
4. Connect some vertices to others by arrows. An arrow may go from a vertex to itself. Each arrow is labelled with one of the numbers $0, 1, \ldots, k$. No two arrows that emanate from the same vertex are allowed to have the same label.
5. One vertex is enclosed in a circle and is called the *initial vertex*.

The resulting graph is called a *diagram*.

Example

See Figure 5.3.

We shall show that every diagram determines a Turing machine whose operation can be described in the following manner. Given a tape and a specific square on the tape, the Turing machine of the initial vertex V of the diagram begins to operate, with its reading head scanning the specified

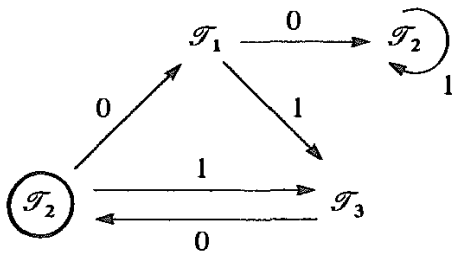


Figure 5.3

square of the tape. If this machine finally stops and the square being scanned at the end of the computation contains the symbol a_i, then we look for an arrow with label i emanating from the vertex V. If there is no such arrow, the computation stops. If there is such an arrow, it leads to a vertex to which another Turing machine has been assigned. Start that machine on the tape produced by the previous computation, at the square that was being scanned at the end of the computation. Repeat the same procedure that was just performed, and keep on doing this until the machine stops. The resulting tape is the output of the machine determined by the diagram. If the machine never stops, then it is not applicable to the initial tape description.

The quadruples for this Turing machine can be specified in the following way.

1. For each occurrence in the diagram of a machine $\mathscr{T}_j$, write its quadruples, changing internal states so that no two machine occurrences have an internal state in common. The initial vertex machine is not to be changed. This retains q_0 as the initial internal state of the machine assigned to the initial vertex. For every other machine occurrence, the original initial state q_0 has been changed to a new internal state.
2. If an occurrence of some $\mathscr{T}_i$ is connected by an arrow $\overset{u}{\rightarrow}$ to some $\mathscr{T}_j$, then, for every (new) internal state q_s of that occurrence of $\mathscr{T}_i$ such that no (new) quadruple of $\mathscr{T}_i$ begins with $q_s a_u$, add the quadruple $q_s a_u a_u q_t$, where q_t is the (new) initial state for $\mathscr{T}_j$. (Step 2 ensures that, whenever $\mathscr{T}_i$ stops while scanning a_u, $\mathscr{T}_j$ will begin operating.)

The following abbreviations are used in diagrams:

1. If one vertex is connected to another vertex by all arrows $\overset{0}{\rightarrow}, \overset{1}{\rightarrow}, \ldots, \overset{k}{\rightarrow}$, we replace the arrows by one unlabelled arrow.
2. If one vertex is connected to another by all arrows except $\overset{u}{\rightarrow}$, we replace all the arrows by $\overset{\neq u}{\rightarrow}$.
3. Let $\mathscr{T}_1\mathscr{T}_2$ stand for $\mathscr{T}_1 \rightarrow \mathscr{T}_2$, let $\mathscr{T}_1\mathscr{T}_2\mathscr{T}_3$ stand for $\mathscr{T}_1 \rightarrow \mathscr{T}_2 \rightarrow \mathscr{T}_3$, and so on. Let $\mathscr{T}^2$ be $\mathscr{T}\mathscr{T}$, let $\mathscr{T}^3$ be $\mathscr{T}\mathscr{T}\mathscr{T}$, and so forth.
4. If no vertex is circled, then the leftmost vertex is to be initial.

To construct diagrams, we need a few simple Turing machines as building blocks.

1. **r** (right machine). Let $\{a_0, a_1, \ldots, a_k\}$ be the alphabet. **r** consists of the quadruples $q_0a_iRq_1$ for all a_i. This machine, which has $k+1$ quadruples, moves one square to the right and then stops.
2. **l** (left machine). Let $\{a_0, a_1, \ldots, a_k\}$ be the alphabet. **l** consists of the quadruples $q_0a_iLq_1$ for all a_i. This machine, which has $k+1$ quadruples, moves one square to the left and then stops.
3. $\mathbf{a_j}$ (constant machine) for the alphabet $\{a_0, a_1, \ldots, a_k\}$. $\mathbf{a_j}$ consists of the quadruples $q_0a_ia_jq_1$ for all a_i. This machine replaces the initial scanned symbol by a_j and then stops. In particular, $\mathbf{a_0}$ erases the scanned symbol, and $\mathbf{a_1}$ prints |.

Examples of Machines Defined by Diagrams

1. **P** (Figure 5.4) finds the first blank to the right of the initially scanned square. In an alphabet $\{a_0, a_1, \ldots, a_k\}$, the quadruples for the machine **P** are: $q_0a_iRq_1$ for all a_i, and $q_1a_ia_iq_0$ for all $a_i \neq a_0$.

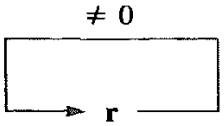


Figure 5.4

2. **Λ** (Figure 5.5) finds the first blank to the left of the initially scanned square.

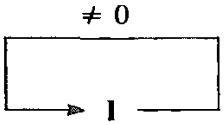


Figure 5.5

Exercises

5.8 Describe the operations of the Turing machines **ρ** (Figure 5.6) and **λ** (Figure 5.7) and write the list of quadruples for each machine.

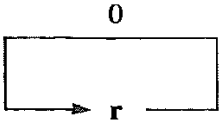


Figure 5.6

Figure 5.7

5.9 Show that machine S in Figure 5.8 searches the tape for a non-blank square. If there are such squares, S finds one and stops. Otherwise, S never stops.

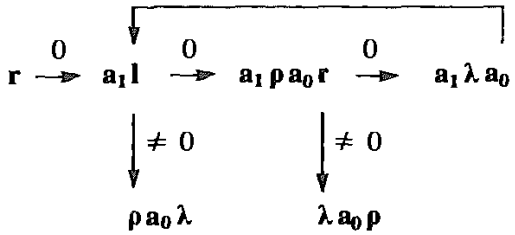


Figure 5.8

To describe some aspects of the operation of a Turing machine on part of a tape, we introduce the following notation:

$\sim$	arbitrary symbol
B...B	sequence of blanks
B...	everything blank to the right
...B	everything blank to the left
W	non-empty word consisting of non-blanks
X	$W_1BW_2B \ldots W_n (n \geq 1)$, a sequence of nonempty words of non-blanks, separated by blanks

Underlining will indicate the scanned symbol.

More Examples of Turing Machines Defined by Diagrams

3. $\mathscr{R}$ (right-end machine). See Figure 5.9.

$$\underline{\sim}\text{XBB} \Rightarrow \sim \text{X}\underline{\text{B}}\text{B}$$

Squares on the rest of the tape are not affected. The same assumption is made in similar places below. When the machine $\mathscr{R}$ begins on a square preceding a sequence of one or more nonempty words, followed by at least two blank squares, it moves right to the first of those blank squares and stops.

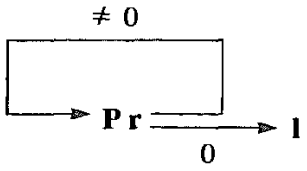


Figure 5.9

4. $\mathscr{L}$ (left-end machine) See Figure 5.10.

$$BBX\underline{\sim} \Rightarrow B\underline{B}X\sim$$

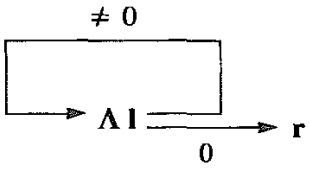


Figure 5.10

5. T (left-translation machine) See Figure 5.11.†

$$\underline{\sim}BWB \Rightarrow \sim W\underline{B}B$$

This machine shifts the whole word W one square to the left.

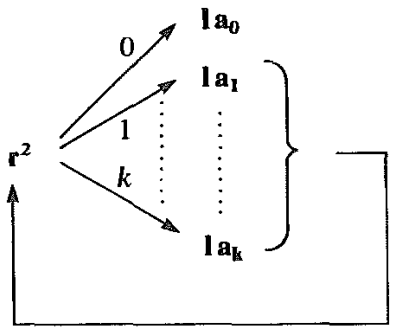


Figure 5.11

6. σ (shift machine). See Figure 5.12.

$$BW_1BW_2\underline{B} \Rightarrow BW_2\underline{B}\ldots B$$

In the indicated situation, W_1 is erased and W_2 is shifted leftward so that it begins where W_1 originally began.

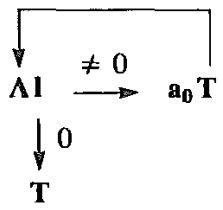


Figure 5.12

7. C (clean-up machine) See Figure 5.13.

$$\sim BBXBW\underline{B} \Rightarrow \sim W\underline{B}\ldots B$$

†There is a separate arrow from $\mathbf{r}^2$ to each of the groups on the right and a separate arrow from each of these, except $\mathbf{1a_0}$, back to $\mathbf{r}^2$.

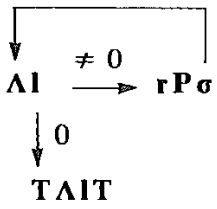


Figure 5.13

8. **K** (word-copier) See Figure 5.14.

$$BW\underline{B}\ldots \Rightarrow BWBW\underline{B}\ldots$$

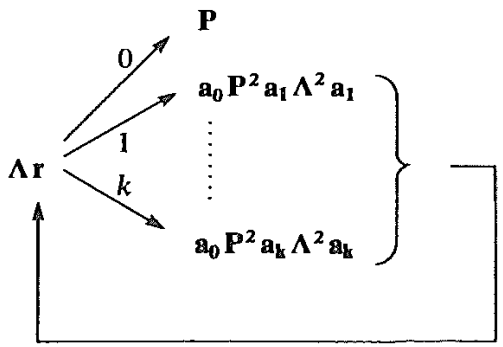


Figure 5.14

9. $\mathbf{K}_n$ (n-shift copier) See Figure 5.15.

$$BW_nBW_{n-1}B\ldots W_1\underline{B}\ldots \Rightarrow BW_nBW_{n-1}B\ldots W_1BW_n\underline{B}\ldots$$

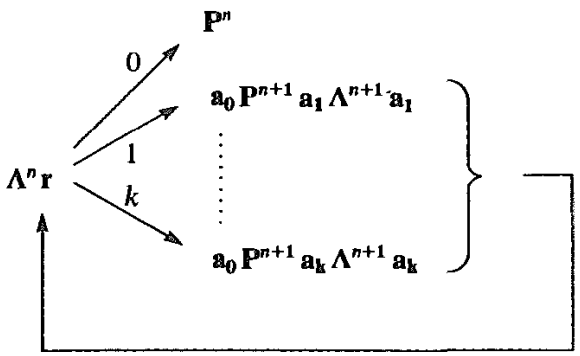


Figure 5.15

Exercises

5.10. Find the number-theoretic function $f(x)$ computed by each of the following Turing machines.

(a) $\mathbf{1a_1}$
(b) Figure 5.16
(c) $\mathbf{PK\Lambda a_1\Lambda(ra_0)^2}$

5.11. Verify that the given functions are computed by the indicated Turing machines.

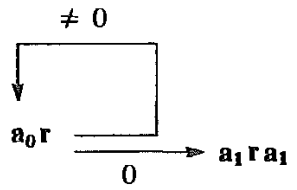


Figure 5.16

(a) $|x - y|$ (Figure 5.17)

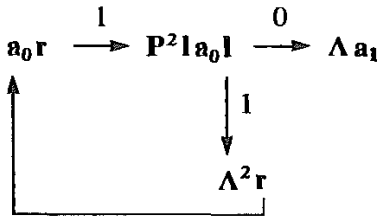


Figure 5.17

(b) $x + y$ $\mathbf{Pa_1\Lambda(ra_0)^2}$
(c) $x \cdot y$ (Figure 5.18)

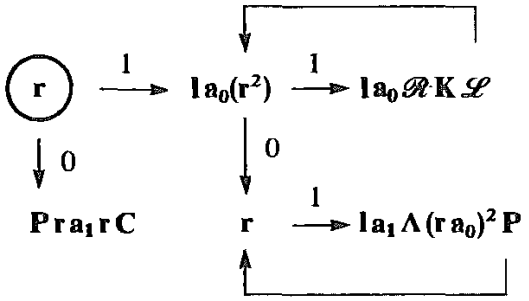


Figure 5.18

5.12. Draw diagrams for Turing machines that will compute the following functions: (a) $\max(x, y)$ (b) $\min(x, y)$ (c) $x \dot{-} y$ (d) $[x/2]$

5.13. Prove that, for any Turing machine $\mathscr{T}$ with alphabet $\{a_0, \ldots, a_k\}$, there is a diagram using the Turing machines $\mathbf{r}, \mathbf{l}, \mathbf{a_0}, \ldots, \mathbf{a_k}$ that defines a Turing machine $\mathscr{S}$ such that $\mathscr{T}$ and $\mathscr{S}$ have the same effect on all tapes. (In fact, $\mathscr{S}$ can be defined so that, except for two additional trivial initial moves left and right, it carries out the same computations as $\mathscr{T}$.)

5.3 PARTIAL RECURSIVE FUNCTIONS. UNSOLVABLE PROBLEMS

Recall, from Section 3.3, that the recursive functions are obtained from the initial functions (the zero function $Z(x)$, the successor function $N(x)$, and the

projection functions $U_i^n(x_1, \ldots, x_n)$) by means of substitution, recursion and the restricted μ-operator. Instead of the restricted μ-operator, let us introduce the *unrestricted μ-operator:*

If $f(x_1, \ldots, x_n) = \mu y(g(x_1, \ldots, x_n, y) = 0)$
$= \text{the least } y \text{ such that } g(x_1, \ldots, x_n, y) = 0$
then f is said to arise from g by means of the unrestricted μ-operator.

Notice that, for some $x_1, \ldots, x_n$, the value $f(x_1, \ldots, x_n)$ need not be defined; this happens when there is no y such that $g(x_1, \ldots, x_n, y) = 0$.

If we replace the restricted μ-operator by the unrestricted μ-operator in the definition of the recursive functions, we obtain a definition of the partial recursive functions. In other words, the *partial recursive functions* are those functions obtained from the initial functions by means of substitution, recursion and the unrestricted μ-operator.

Whereas all recursive functions are total functions, some partial recursive functions will not be total functions. For example, $\mu y(x + y = 0)$ is defined only when $x = 0$.

Since partial recursive functions may not be defined for certain arguments, the definition of the unrestricted μ-operator should be made more precise:

$\mu y(g(x_1, \ldots, x_n, y) = 0) = k$ means that, for $0 \leqslant u < k$,
$g(x_1, \ldots, x_n, u)$ is defined and $g(x_1, \ldots, x_n, u) \neq 0$, and
$g(x_1, \ldots, x_n, y) = 0$.

Observe that, if $R(x_1, \ldots, x_n, y)$ is a recursive relation, then $\mu y(R(x_1, \ldots, x_n, y))$ can be considered an admissible application of the unrestricted μ-operator. In fact, $\mu y(R(x_1, \ldots, x_n, y)) = \mu y(C_R(x_1, \ldots, x_n, y) = 0)$, where C_R is the characteristic function of R. Since R is a recursive relation, C_R is, by definition, a recursive function.

Exercises

5.14 Describe the following partial recursive functions.

(a) $\mu y(x + y + 1 = 0)$
(b) $\mu y(y > x)$
(c) $\mu y(y + x = x)$

5.15 Show that all recursive functions are partial recursive.
5.16 Show that every partial function whose domain is a finite set of natural numbers is a partial recursive function.

It is easy to convince ourselves that every partial recursive function $f(x_1, \ldots, x_n)$ is computable, in the sense that there is an algorithm that computes $f(x_1, \ldots, x_n)$ when $f(x_1, \ldots, x_n)$ is defined and gives no result when $f(x_1, \ldots, x_n)$ is undefined. This property is clear for the initial functions and

is inherited under the operations of substitution, recursion and the unrestricted μ-operator.

It turns out that the partial recursive functions are identical with the Turing-computable functions. To show this, it is convenient to introduce a different kind of Turing-computablility.

A partial number-theoretic function $f(x_1, \ldots, x_n)$ is said to be *standard Turing-computable* if there is a Turing machine $\mathscr{T}$ such that, for any natural numbers $k_1, \ldots, k_n$, the following holds.

> Let $B\overline{k_1}B\overline{k_2}B \ldots B\overline{k_n}$ be called the *argument strip*.† Notice that the argument strip is $B\overline{(k_1, \ldots, k_n)}$. Take any tape containing the argument strip but without any symbols to the right of it. (It may contain symbols to the left.) The machine $\mathscr{T}$ is begun on this tape with its reading head scanning the first | of $\overline{k_1}$. Then:

1. $\mathscr{T}$ stops if and only if $f(k_1, \ldots, k_n)$ is defined.
2. If $\mathscr{T}$ stops, the tape contains the same argument strip as before, followed by $B\overline{f(k_1, \ldots k_n)}$. Thus, the final tape contains

$$B\overline{k_1}B\overline{k_2}B \ldots B\overline{k_n}B\overline{f(k_1, \ldots, k_n)}$$

Moreover:

3. The reading head is scanning the first | of $\overline{f(k_1, \ldots, k_n)}$.
4. There is no non-blank symbol on the tape to the right of $\overline{f(k_1, \ldots, k_n)}$.
5. During the entire computation, the reading head never scans any square to the left of the argument strip.

For the sake of brevity, we shall say that the machine $\mathscr{T}$ described above *ST-computes* the function $f(x_1, \ldots, x_n)$.

Thus, the additional requirement of standard Turing computability is that the original arguments are preserved, the machine stops if and only if the function is defined for the given arguments, and the machine operates on or to the right of the argument strip. In particular, anything to the left of the argument strip remains unchanged.

PROPOSITION 5.1

Every standard Turing-computable function is Turing-computable.

Proof

Let $\mathscr{T}$ be a Turing machine that ST-computes a partial function $f(x_1, \ldots, x_n)$. Then f is Turing-computable by the Turing machine $\mathscr{T}\mathbf{PC}$. In

†For a function of one variable, the argument strip is taken to be $B\overline{k_1}$.

fact, after $\mathscr{T}$ operates, we obtain $\mathrm{B}\overline{x_1}\mathrm{B}\ldots\mathrm{B}\overline{x_n}\mathrm{B}\overline{f(x_1,\ldots,x_n)}$, with the reading head at the leftmost | of $\overline{f(x_1,\ldots,x_n)}$. **P** then moves the reading head to the right of $\overline{f(x_1,\ldots,x_n)}$, and then **C** removes the original argument strip.

PROPOSITION 5.2

Every partial recursive function is standard Turing-computable.

Proof

(a) $\mathbf{Pra_1}$ ST-computes the zero function $Z(x)$.

(b) The successor function $N(x)$ is ST-computed by $\mathbf{PKa_1Ar}$.

(c) The projection function $U_i^n(x_1,\ldots,x_n)=x_i$ is ST-computed by $\mathscr{R}\mathbf{K}_{n-i+1}\mathbf{Ar}$.

(d) (*Substitution.*) Let $f(x_1,\ldots,x_n)=g(h_1(x_1,\ldots,x_n),\ldots,h_m(x_1,\ldots,x_n))$ and assume that $\mathscr{T}$ ST-computes g and $\mathscr{T}_j$ ST-computes h_j for $1\leqslant j\leqslant m$. Let $\mathscr{S}_j$ be the machine $\mathscr{T}_j\mathbf{P}\sigma^n(\mathbf{K}_{n+j})^n\Lambda^n\mathbf{r}$. The reader should verify that f is ST-computed by

$$\mathscr{T}_1\mathbf{P}(\mathbf{K}_{n+1})^n\Lambda^n\mathbf{r}\mathscr{S}_2\mathscr{S}_3\ldots\mathscr{S}_{m-1}\mathscr{T}_m\mathbf{P}\sigma^n\Lambda^m\mathbf{r}\mathscr{T}\sigma'^n\mathbf{Ar}$$

We take advantage of the ST-computability when, storing $\overline{x_1},\ldots,\overline{x_n}$, $\overline{h_1(x_1,\ldots,x_n)},\ldots,\overline{h_i(x_1,\ldots,x_n)}$ on the tape, we place $\overline{(x_1,\ldots,x_n)}$ on the tape to the right and compute $\overline{h_{i+1}(x_1,\ldots,x_n)}$ without disturbing what we have stored on the left.

(e) (*Recursion.*) Let

$$f(x_1,\ldots,x_n,0)=g(x_1,\ldots,x_n)$$
$$f(x_1,\ldots,x_n,y+1)=h(x_1,\ldots,x_n,y,f(x_1,\ldots,x_n,y))$$

Assume that $\mathscr{S}$ ST-computes g and $\mathscr{T}$ ST-computes h. Then the reader should verify that the machine in Figure 5.19 ST-computes f.

(f) *Unrestricted μ-operator.* Let $f(x_1,\ldots x_n)=\mu y(g(x_1,\ldots,x_n,y)=0)$ and assume that $\mathscr{T}$ ST-computes g. Then the machine in Figure 5.20 ST-computes f.

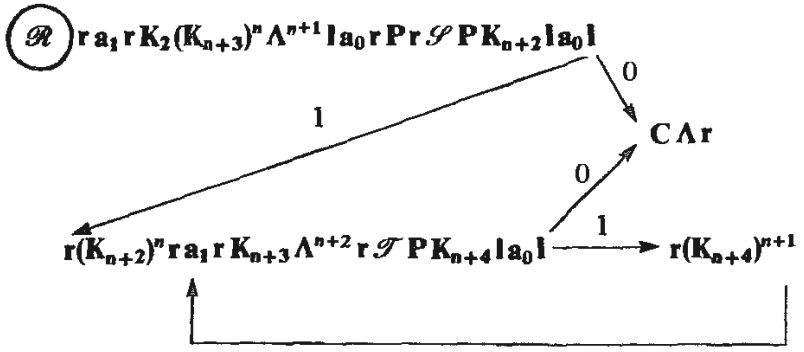


Figure 5.19

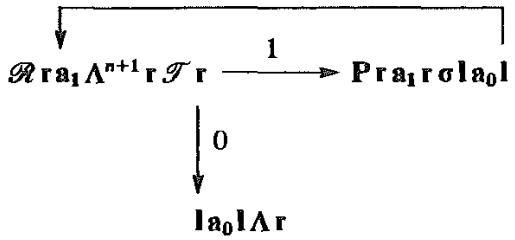


Figure 5.20

Exercise

5.17 For a recursion of the form

$$f(0) = k$$
$$f(y+1) = h(y, f(y))$$

show how the diagram in Figure 5.19 must be modified.

COROLLARY 5.3

Every partial recursive function is Turing-computable.

Exercise

5.18 Prove that every partial recursive function is Turing-computable by a Turing machine with alphabet $\{a_0, a_1\}$.

In order to prove the converse of Corollary 5.3, we must arithmetize the language of Turing computability by assigning numbers, called *Gödel numbers*, to the expressions arising in our study of Turing machines. 'R' and 'L' are assigned the Gödel numbers 3 and 5, respectively. The tape symbols a_i are assigned the numbers $7 + 4i$, while the internal state symbols q_i are given the numbers $9 + 4i$. For example, the blank B, which is a_0, receives the number 7; the stroke |, which is a_1, has the number 11; and the initial internal state symbol q_0 has the number 9. Notice that all symbols have odd Gödel numbers, and different symbols have different numbers assigned to them.

As in Section 3.4, a finite sequence $u_0, u_1, \ldots, u_k$ of symbols is assigned the Gödel number $p_0^{g(u_0)} p_1^{g(u_1)} \ldots p_k^{g(u_k)}$, where p_0, p_1, $p_2, \ldots$ are the prime numbers $2, 3, 5, \ldots$ in ascending order and $g(u_i)$ is the Gödel number assigned to u_i. For example, the quadruple $q_0 a_0 a_1 q_0$ receives the Gödel number $2^9 3^7 5^{11} 7^9$.

By an *expression* we mean a finite sequence of symbols. We have just shown how to assign Gödel numbers to expressions. In a similar manner, to any finite sequence $E_0, E_1, \ldots, E_m$ of expressions we assign the number

$p_0^{g(E_0)} p_1^{g(E_1)} \ldots p_m^{g(E_m)}$. For example, this assigns Gödel numbers to finite sequences of Turing machine quadruples and to finite sequences of tape descriptions. Observe that the Gödel number of an expression is even and, therefore, different from the Gödel number of a symbol, which is odd. Moreover, the Gödel number of a sequence of expressions has an even number as an exponent of p_0 and is, therefore, different from the Gödel number of an expression, which has an odd number as an exponent of p_0.

The reader should review Sections 3.3 and 3.4, especially the functions $\ell h(x), (x)_i)$, and $x * y$. Assume that x is the Gödel number of a finite sequence $W_0, W_1, \ldots, W_k$; that is, $x = p_0^{g(W_0)} p_1^{g(W_1)} \ldots p_k^{g(W_k)}$, where $g(W_j)$ is the Gödel number of W_j. Recall that $\ell h(x) = k + 1$, the length of the sequence, and $(x)_j = g(W_j)$, the Gödel number of the jth term of the sequence. If in addition, y is the Gödel number of a finite sequence $V_0, V_1, \ldots, V_m$, then $x * y$ is the Gödel number of the juxtaposition $W_0, W_1, \ldots, W_k, V_0, V_1, \ldots, V_m$ of the two sequences.

PROPOSITION 5.4

The following number-theoretic relations and functions are primitive recursive. In each case, we write first the notation for the relation or function, then, the intuitive interpretation in terms of Turing machines, and, finally, the exact definition. (For the proofs of primitive recursiveness, use Proposition 3.18 and various primitive relations and functions defined in Section 3.3. At a first reading, it may be advisable to concentrate on just the intuitive meanings and postpone the technical verification until later.)

IS(x): x is the Gödel number of an internal state symbol q_u:

$$(\exists u)_{u<x}(x = 9 + 4u)$$

Sym(x): x is the Gödel number of an alphabet symbol a_u:

$$(\exists u)_{u<x}(x = 7 + 4u)$$

Quad(x): x is the Gödel number of a Turing machine quadruple:

$$\ell h(x) = 4 \wedge \mathrm{IS}((x)_0) \wedge \mathrm{Sym}((x)_1) \wedge \mathrm{IS}((x)_3) \wedge [\mathrm{Sym}((x)_2) \vee (x)_2 = 3 \vee (x)_2 = 5]$$

TM(x): x is the Gödel number of a Turing machine (in the form of a finite sequence of appropriate quadruples):

$$(\forall u)_{u<\ell h(x)}\mathrm{Quad}((x)_u) \wedge x > 1 \wedge (\forall u)_{u<\ell h(x)}(\forall v)_{v<\ell h(x)}(u \neq v \Rightarrow [((x)_u)_0 \neq ((x)_v)_0 \vee ((x)_u)_1 \neq ((x)_v)_1])$$

TD(x): x is the Gödel number of a tape description:

$$x > 1 \wedge (\forall u)_{u<\ell h(x)}[\mathrm{IS}((x)_u) \vee \mathrm{Sym}((x)_u)] \wedge (\exists_1 u)_{u<\ell h(x)}\mathrm{IS}((x)_u) \wedge (\forall u)_{u<\ell h(x)}(\mathrm{IS}((x)_u) \Rightarrow u + 1 < \ell h(x))$$

Cons(x, y, z): x and y are Gödel numbers of tape descriptions α and β, and z is the Gödel number of a Turing machine quadruple that transforms α into β:

$$\begin{aligned}
&\mathrm{TD}(x) \wedge \mathrm{TD}(y) \wedge \mathrm{Quad}(z) \wedge (\exists w)_{w<\ell h(x)\dot{-}1}[\mathrm{IS}((x)_w)\\
&\wedge (x)_w = (z)_0 \wedge (x)_{w+1} = (z)_1 \wedge\\
&\mathrm{I}\begin{cases}([\mathrm{Sym}((z)_2) \wedge (y)_{w+1} = (z)_2 \wedge (y)_w = (z)_3 \wedge \ell h(x) = \ell h(y)\\ \wedge(\forall u)_{u<\ell h(x)}(u \neq w \wedge u \neq w+1 \Rightarrow (x)_u = (y)_u)]\vee\end{cases}\\
&\mathrm{II}\begin{cases}[(z)_2 = 3 \wedge (y)_w = (x)_{w+1} \wedge (y)_{w+1} = (z)_3 \wedge\\ (\forall u)_{u<\ell h(x)}(u \neq w \wedge u \neq w+1 \Rightarrow (y)_u = (x)_u) \wedge\\ ([w+2 < \ell h(x) \wedge \ell h(y) = \ell h(x)] \vee [w+2 = \ell h(x) \wedge\\ \ell h(y) = \ell h(x)+1 \wedge (y)_{w+2} = 7])]\vee\end{cases}\\
&\mathrm{III}\begin{cases}[(z)_2 = 5 \wedge \{[w \neq 0 \wedge (y)_{w\dot{-}1} = (z)_3 \wedge (y)_w = (x)_{w\dot{-}1}\\ \wedge\ \ell h(y) = \ell h(x) \wedge (\forall u)_{u<\ell h(x)}(u \neq w\dot{-}1 \wedge u \neq w \Rightarrow\\ (y)_u = (x)_u)] \vee [w = 0 \wedge (y)_0 = (z)_3 \wedge (y)_1 = 7 \wedge\\ \ell h(y) = \ell h(x)+1 \wedge (\forall u)_{0<u<\ell h(x)}(y)_{u+1} = (x)_u]\}])]\end{cases}
\end{aligned}$$

I corresponds to a quadruple $q_j a_i a_k q_r$, II to a quadruple $q_j a_i R q_r$, and III to a quadruple $q_j a_i L q_r$.

NTD(x): x is the Gödel number of a numerical tape description – that is, a tape description in which the tape has the form $E_1 \bar{k} E_2$, where each of E_1 and E_2 is empty or consists entirely of blanks, and the location of the reading head is arbitrary:

$$\begin{aligned}
&\mathrm{TD}(x) \wedge (\forall u)_{u<\ell h(x)}(\mathrm{Sym}((x)_u) \Rightarrow (x)_u = 7 \vee (x)_u = 11)\\
&\wedge (\forall u)_{u<\ell h(x)}(\forall v)_{v<\ell h(x)}(\forall w)_{w<\ell h(x)}(u < v \wedge v < w \wedge (x)_u = 11 \wedge\\
&(x)_w = 11 \Rightarrow (x)_v \neq 7)\ (\exists u)_{u<\ell h(x)}((x)_u = 11)
\end{aligned}$$

Stop(x, z): z is the Gödel number of a Turing machine $\mathscr{T}$ and x is the Gödel number of a tape description α such that $\mathscr{T}$ stops at α:

$$\begin{aligned}
\mathrm{TM}(z) \wedge \mathrm{TD}(x) \wedge \neg(\exists u)_{u<\ell h(x)}[&\mathrm{IS}((x)_u) \wedge (\exists v)_{v<\ell h(z)}(((z)_v)_0\\
&= (x)_u \wedge ((z)_v)_1 = (x)_{u+1})]
\end{aligned}$$

Comp(y, z): z is the Gödel number of a Turing machine $\mathscr{T}$ and y is the Gödel number of a computation of $\mathscr{T}$:

$$\begin{aligned}
&y > 1 \wedge \mathrm{TM}(z) \wedge (\forall u)_{u<\ell h(y)}\mathrm{TD}((y)_u) \wedge \mathrm{Stop}((y)_{\ell h(y)\dot{-}1}, z) \wedge\\
&(\forall u)_{u<\ell h(y)\dot{-}1}(\exists w)_{w<\ell h(z)}\mathrm{Cons}((y)_u, (y)_{u+1}, (z)_w) \wedge\\
&(\forall v)_{v<\ell h((y)_0)}(\mathrm{IS}(((y)_0)_v) \Rightarrow ((y)_0)_v = 9)
\end{aligned}$$

Num(x): The Gödel number of the word $\bar{x}$ – that is, of $|^{x+1}$:

$$\mathrm{Num}(x) = \prod_{u \leqslant x} p_u^{11}$$

$\mathrm{TR}(x_1,\ldots,x_n)$: The Gödel number of the tape representation $\overline{(x_1,\ldots,x_n)}$ of the n-tuple $(x_1,\ldots,x_n)$:

$$\mathrm{TR}(x_1,\ldots,x_n) = \mathrm{Num}(x_1) * 2^7 * \mathrm{Num}(x_2) * 2^7 * \cdots * 2^7 * \mathrm{Num}(x_n)$$

$U(y)$: If y is the Gödel number of a computation that results in a numerical tape description, then $U(y)$ is the number represented on that final tape.†

$$U(y) = \left[\sum_{u<\ell h((y)_{\ell h(y)\dot{-}1})} \overline{\mathrm{sg}}(|((y)_{\ell h(y)\dot{-}1})_u - 11|)\right] \dot{-} 1$$

[Let w be the number, represented by $|^{w+1}$, on the final tape. The calcualtion of $U(y)$ tallies a 1 for every stroke | that appears on the final tape. This yields a sum of $w+1$, and then 1 is subtracted to obtain w.]

$T_n(z,x_1,\ldots,x_n,y)$: y is the Gödel number of a computation of a Turing machine with Gödel number z such that the computation begins on the tape $(\overline{x_1,\ldots,x_n})$, with the reading head scanning the first | in $\bar{x}_1$, and ends with a numerical tape description:

$$\mathrm{Comp}(y,z) \wedge (y)_0 = 2^9 * \mathrm{TR}(x_1,\ldots,x_n) \wedge \mathrm{NTD}((y)_{\ell h(y)\dot{-}1})$$

When $n=1$, replace $\mathrm{TR}(x_1,\ldots,x_n)$ by $\mathrm{Num}(x_1)$. (Observe that, if $T_n(z,x_1,\ldots,x_n,y_1)$ and $T_n(z,x_1,\ldots,x_n,y_2)$, then $y_1 = y_2$, since there is at most one computation of a Turing machine starting with a given initial tape.)

PROPOSITION 5.5

If $\mathscr{T}$ is a Turing machine that computes a number-theoretic function $f(x_1,\ldots,x_n)$ and e is a Gödel number of $\mathscr{T}$, then‡

$$f(x_1,\ldots,x_n) = U(\mu y T_n(e,x_1,\ldots,x_n,y))$$

Proof

Let $k_1,\ldots,k_n$ be any natural numbers. Then $f(k_1,\ldots,k_n)$ is defined if and only if there is a computation of $\mathscr{T}$ beginning with $(\overline{k_1,\ldots,k_n})$ and ending with a numerical tape description – that is, if and only if $(\exists y)T_n(e,k_1,\ldots,x_n,y)$. So, $f(k_1,\ldots,k_n)$ is defined if and only if $\mu y T_n(e,k_1,\ldots,k_n,y)$ is defined. Moreover, when $f(k_1,\ldots,k_n)$ is defined,

†If y is not the Gödel number of a computation that yields a numerical tape description, $U(y)$ is defined, but its value in such cases will be of no significance.

‡Remember that an equality between two partial functions means that, whenever one of them is defined, the other is also defined and the two functions have the same value.

$\mu y T_n(e, k_1, \ldots, k_n, y)$ is the Gödel number of a computation of $\mathscr{T}$ beginning with $(k_1, \ldots, k_n)$ and $U(\mu y T_n(e, k_1, \ldots, k_n, y))$ is the value yielded by the computation, namely, $f(k_1, \ldots, k_n)$.

COROLLARY 5.6

Every Turing-computable function is partial recursive.

Proof

Assume $f(x_1, \ldots, x_n)$ is Turing-computable by a Turing machine with Gödel number e. Then $f(x_1, \ldots, x_n) = U(\mu y T_n(e, x_1, \ldots, x_n, y))$. Since T_n is primitive recursive, $\mu y T_n(e, x_1, \ldots, x_n, y)$ is partial recursive. Hence, $U(\mu y T_n(e, x_1, \ldots, x_n, y))$ is partial recursive.

COROLLARY 5.7

The Turing-computable functions are identical with the partial recursive functions.

Proof

Use Corollaries 5.6 and 5.3.

COROLLARY 5.8

Every total partial recursive function is recursive.

Proof

Assume that the total partial recursive function $f(x_1, \ldots, x_n)$ is Turing-computable by the Turing machine with Gödel number e. Then, for all $x_1, \ldots, x_n$, $(\exists y) T_n(e, x_1, \ldots, x_n, y)$. Hence, $\mu y T_n(e, x_1, \ldots, x_n, y)$ is produced by an application of the restricted μ-operator and is, therefore, recursive. So, $U(\mu y T_n(e, x_1, \ldots, x_n, y))$ is also recursive. Now use Proposition 5.5.

COROLLARY 5.9

For any total number-theoretic function f, f is recursive if and only if f is Turing-computable.

Proof

Use Corollaries 5.7–5.8 and Exercise 5.15.

Church's thesis amounts to the assertion that the recursive functions are the same as the computable total functions. By Corollary 5.9, this is equivalent to the identity, for total functions, of computability and Turing computability. This strengthens the case for Church's thesis because of the plausibility of the identification of Turing computability with computability. Let us now widen Church's thesis to assert that the computable functions (partial or total) are the same as the Turing-computable functions. By Corollary 5.7, this implies that a function is computable if and only if it is partial recursive.

COROLLARY 5.10

Any number-theoretic function is Turing-computable if and only if it is standard Turing-computable.

Proof

Use Proposition 5.1, Corollary 5.6 and Proposition 5.2.

COROLLARY 5.11 (KLEENE'S NORMAL FORM THEOREM)

As z varies over all natural numbers, $U(\mu y T_n(z, x_1, \ldots, x_n, y))$ enumerates with repetitions all partial recursive functions of n variables.

Proof

Use Corollary 5.3 and Proposition 5.5. The fact that every partial recursive function of n variables reappears for infinitely many z follows from Exercise 5.7. (Notice that, when z is not the Gödel number of a Turing machine, there is no y such that $T_n(z, x_1, \ldots, x_n, y)$, and, therefore, the corresponding partial recursive function is the empty function.†)

COROLLARY 5.12

For any recursive relation $R(x_1, \ldots, x_n, y)$, there exist natural numbers z_0 and v_0 such that, for all natural numbers $x_1, \ldots, x_n$:

†The empty function is the empty set $\emptyset$. It has the empty set as its domain.

(a) $(\exists y)R(x_1,\ldots,x_n,y)$ if and only if $(\exists y)T_n(z_0,x_1,\ldots,x_n,y)$
(b) $(\forall y)R(x_1,\ldots,x_n,y)$ if and only if $(\forall y)\neg T_n(v_0,x_1,\ldots,x_n,y)$

Proof

(a) The function $f(x_1,\ldots,x_n)=\mu y R(x_1,\ldots,x_n,y)$ is partial recursive. Let z_0 be a Gödel number of a Turing machine that computes f. Hence, $f(x_1,\ldots,x_n)$ is defined if and only if $(\exists y)T_n(z_0,x_1,\ldots,x_n,y)$. But $f(x_1,\ldots,x_n)$ is defined if and only if $(\exists y)R(x_1,\ldots,x_n,y)$.
(b) Applying part (a) to the recursive relation $\neg R(x_1,\ldots,x_n,y)$, we obtain a number v_0 such that:

$$(\exists y)\neg R(x_1,\ldots,x_n,y) \text{ if and only if } (\exists y)T_n(v_0,x_1,\ldots,x_n,y)$$

Now take the negations of both sides of this equivalence.

Exercise

5.19 Extend Corollary 5.12 to two or more quantifiers. For example, if $R(x_1,\ldots,x_n,y,z)$ is a recursive relation, show that there are natural numbers z_0 and v_0 such that, for all $x_1,\ldots,x_n$:
(a) $(\forall z)(\exists y)R(x_1,\ldots,x_n,y,z)$ if and only if $(\forall z)(\exists y)T_{n+1}(z_0,x_1,\ldots,x_n,y,z)$.
(b) $(\exists z)(\forall y)R(x_1,\ldots,x_n,y,z)$ if and only if $(\exists z)(\forall y)\neg T_{n+1}(v_0,x_1,\ldots,x_n,y,z)$.

COROLLARY 5.13

(a) $(\exists y)T_n(x_1,x_1,x_2,\ldots,x_n,y)$ is not recursive.
(b) $(\exists y)T_n(z,x_1,\ldots,x_n,y)$ is not recursive.

Proof

(a) Assume $(\exists y)T_n(x_1,x_1,x_2,\ldots,x_n,y)$ is recursive. Then the relation $\neg(\exists y)T_n(x_1,x_1,x_2,\ldots,x_n,y)\wedge z=z$ is recursive. So, by Corollary 5.12(a), there exists z_0 such that:

$$(\exists z)(\neg(\exists y)T_n(x_1,x_1,x_2,\ldots,x_n,y)\wedge z=z) \text{ if and only if}$$
$$(\exists z)T_n(z_0,x_1,x_2,\ldots,x_n,z)$$

Hence, since z obviously can be omitted on the left,

$$\neg(\exists y)T_n(x_1,x_1,x_2,\ldots,x_n,y) \text{ if and only if } (\exists z)T_n(z_0,x_1,x_2,\ldots,x_n,z)$$

Let $x_1=x_2=\cdots=x_n=z_0$. Then we obtain the contradiction

$$\neg(\exists y)T_n(z_0,z_0,z_0,\ldots,z_0,y) \text{ if and only if } (\exists z)T_n(z_0,z_0,z_0,\ldots,z_0,z)$$

(b) If $(\exists y)T_n(z,x_1,x_2,\ldots,x_n,y)$ were recursive, so would be, by substitution, $(\exists y)T_n(x_1,x_1,x_2,\ldots,x_n,y)$, contradicting part (a).

Exercises

5.20 Prove that there is a partial recursive function $g(z,x)$ such that, for any partial recursive function $f(x)$, there is a number z_0 for which $f(x) = g(z_0,x)$ holds for all x. Then show that there must exist a number v_0 such that $g(v_0, v_0)$ is not defined.

5.21 Let $h_1(x_1,\ldots,x_n),\ldots,h_k(x_1,\ldots,x_n)$ be partial recursive functions, and let $R_1(x_1,\ldots,x_n),\ldots,R_k(x_1,\ldots,x_n)$ be recursive relations that are exhaustive (i.e., for any $x_1,\ldots,x_n$, at least one of the relations holds) and pairwise mutually exclusive (i.e., for any $x_1,\ldots,x_n$, no two of the relations hold). Define

$$g(x_1,\ldots,x_n) = \begin{cases} h_1(x_1,\ldots,x_n) & \text{if } R_1(x_1,\ldots,x_n) \\ \ldots\ldots\ldots & \\ h_k(x_1,\ldots,x_n) & \text{if } R_k(x_1,\ldots,x_n) \end{cases}$$

Prove that g is partial recursive.

5.22 A partial function $f(x)$ is said to be recursively completable if there is a recursive function $h(x)$ such that, for every x in the domain of f, $h(x) = f(x)$.

(a) Prove that $\mu y T_1(x,x,y)$ is not recursively completable.
(b) Prove that a partial recursive function $f(x)$ is recursively completable if the domain D of f is a recursive set – that is, if the property '$x \in D$' is recursive.
(c) Find a partial recursive function $f(x)$ that is recursively completable but whose domain is not recursive.

5.23 If $R(x,y)$ is a recursive relation, prove that there are natural numbers z_0 and v_0 such that:

(a) $\neg[(\exists y)R(z_0,y) \Leftrightarrow (\forall y)\neg T_1(z_0,z_0,y)]$
(b) $\neg[(\forall y)R(v_0,y) \Leftrightarrow (\exists y)T_1(v_0,v_0,y)]$

5.24 If $S(x)$ is a recursive property, show that there are natural numbers z_0 and v_0 such that:

(a) $\neg[S(z_0) \Leftrightarrow (\forall y)\neg T_1(z_0,z_0,y)]$
(b) $\neg[S(v_0) \Leftrightarrow (\exists y)T_1(v_0,v_0,y)]$

5.25 Show that there is no recursive function $B(z,x_1,\ldots,x_n)$ such that, if z is a Gödel number of a Turing number $\mathscr{T}$ and $k_1,\ldots,k_n$ are natural numbers for which $f_{\mathscr{T},n}(k_1,\ldots,k_n)$ is defined, then the number of steps in the computation of $f_{\mathscr{T},n}(k_1,\ldots,k_n)$ is less than $B(z,k_1,\ldots,k_n)$.

Let $\mathscr{T}$ be a Turing machine. The *halting problem* for $\mathscr{T}$ is the problem of determining, for each tape description β, whether $\mathscr{T}$ is applicable to β, that is, whether there is a computation of $\mathscr{T}$ that begins with β.

We say that the halting problem for $\mathscr{T}$ is *algorithmically solvable* if there is an algorithm that, given a tape description β, determines whether $\mathscr{T}$ is

applicable to β. Instead of a tape description β, we may assume that the algorithm is given the Gödel number of β. Then the desired algorithm will be a computable function $H_{\mathscr{T}}$ such that:

$$H_{\mathscr{T}}(x) = \begin{cases} 0 & \text{if } x \text{ is the Gödel number of a tape description } \beta \\ & \text{to which } \mathscr{T} \text{ is applicable} \\ 1 & \text{otherwise} \end{cases}$$

If we accept Turing algorithms as exact counterparts of algorithms (that is, the extended Church's thesis), then the halting problem for $\mathscr{T}$ is algorithmically solvable if and only if the function $H_{\mathscr{T}}$ is Turing-computable, or equivalently, by Corollary 5.9, recursive. When the function $H_{\mathscr{T}}$ is recursive, we say that the halting problem for $\mathscr{T}$ is *recursively solvable*. If $H_{\mathscr{T}}$ is not recursive, we say that the halting problem for $\mathscr{T}$ is *recursively unsolvable*.

PROPOSITION 5.14

There is a Turing machine with a recursively unsolvable halting problem.

Proof

By Proposition 5.2, let $\mathscr{T}$ be a Turing machine that ST-computes the partial recursive function $\mu y T_1(x,x,y)$. Remember that, by the definition of standard Turing computability, if $\mathscr{T}$ is begun on the tape consisting of only $\bar{x}$ with its reading head scanning the leftmost |, then $\mathscr{T}$ stops if and only if $\mu y T_1(x,x,y)$ is defined. Assume that $\mathscr{T}$ has a recursively solvable halting problem, that is, that the function $H_{\mathscr{T}}$ is recursive. Recall that the Gödel number of the tape description $q_0\bar{x}$ is $2^9 * \text{Num}(x)$. Now,

$$\begin{aligned} (\exists y)T_1(x,x,y) \text{ if and only if } & \quad \mu y T_1(x,x,y) \text{ is defined} \\ \text{if and only if } & \quad \mathscr{T}, \text{ begun on } q_0\bar{x}, \text{ performs a computation} \\ \text{if and only if } & \quad H_{\mathscr{T}}(2^9 * \text{Num}(x)) = 0 \end{aligned}$$

Since $H_{\mathscr{T}}$, Num and $*$ are recursive, $(\exists y)T_1(x,x,y)$ is recursive, contradicting Corollary 5.13(a) (when $n = 1$).

Exercises

5.26 Give an example of a Turing machine with a recursively solvable halting problem.

5.27 Show that the following *special halting problem* is recursively unsolvable: given a Gödel number z of a Turing machine $\mathscr{T}$ and a natural number x, determine whether $\mathscr{T}$ is applicable to $q_0\bar{x}$.

5.28 Show that the following *self-halting problem* is recursively unsolvable: given a Gödel number z of a Turing machine $\mathscr{T}$, determine whether $\mathscr{T}$ is applicable to $q_0\bar{z}$.

5.29 The *printing problem* for a Turing machine $\mathscr{T}$ and a symbol a_k is the problem of determining, for any given tape description α, whether $\mathscr{T}$, begun on α, ever prints the symbol a_k. Find a Turing machine $\mathscr{T}$ and a symbol a_k for which the printing problem is recursively unsolvable.

5.30 Show that the following decision problem is recursively unsolvable: given any Turing machine $\mathscr{T}$, if $\mathscr{T}$ is begun on an empty tape, determine whether $\mathscr{T}$ stops (that is, whether $\mathscr{T}$ is applicable to q_0B).

5.31[D] Show that the problem of deciding, for any given Turing machine, whether it has a recursively unsolvable halting problem is itself recursively unsolvable.

To deal with more intricate decision problems and other aspects of the theory of computability, we need more powerful tools. First of all, let us introduce the notation

$$\varphi_z^n(x_1, \ldots, x_n) = U(\mu y T_n(z, x_1, \ldots, x_n, y))$$

Thus, by Corollary 5.11, $\varphi_0^n, \varphi_1^n, \varphi_2^n, \ldots$ is an enumeration of all partial recursive functions of n variables. The subscript j is called an *index* of the function φ_j^n. Each partial recursive function of n variables has infinitely many indices.

PROPOSITION 5.15 (ITERATION THEOREM OR *s-m-n* THEOREM)

For any positive integers m and n, there is a primitive recursive function $s_n^m(z, y_1, \ldots, y_m)$ such that

$$\varphi_z^{m+n}(x_1, \ldots, x_n, y_1, \ldots, y_m) = \varphi_{s_n^m(z, y_1, \ldots, y_m)}(x_1, \ldots, x_n)$$

Thus, not only does assigning particular values to $z, y_1, \ldots, y_m$ in $\varphi_z^{m+n}(x_1, \ldots, x_n, y_1, \ldots, y_m)$ yield a new partial recursive function of n variables, but also the index of the resulting function is a primitive recursive function of the old index z and of $y_1, \ldots, y_m$.

Proof

If $\mathscr{T}$ is a Turing machine with Gödel number z, let $\mathscr{T}_{y_1, \ldots, y_m}$ be a Turing machine that, when begun on $\overline{(x_1, \ldots, x_n)}$, produces $\overline{(x_1, \ldots, x_n, y_1, \ldots, y_m)}$, moves back to the leftmost | of $\overline{x_1}$, and then behaves like $\mathscr{T}$. Such a machine is defined by the diagram

$$\mathbf{R}\mathbf{r}(\mathbf{a}_1\mathbf{r})^{y_1+1}\mathbf{r}(\mathbf{a}_1\mathbf{r})^{y_2+1}\mathbf{r} \ldots \mathbf{r}(\mathbf{a}_1\mathbf{r})^{y_m+1}\mathscr{L}\mathbf{r}\mathscr{T}$$

The Gödel number $s_n^m(z, y_1, \ldots, y_m)$ of this Turing machine can be effectively computed and, by Church's thesis, would be partial recursive. In fact, s_n^m can be computed by a primitive recursive function $g(z, y_1, \ldots, y_m)$ defined in the following manner. Let $t = y_1 + \ldots + y_m + 2m + 1$. Also, let $u(i) =$

$2^{9+4i}3^7 5^{11} 7^{9+4i}$ and $v(i) = 2^{9+4i}3^{11}5^3 7^{13+4i}$. Notice that $u(i)$ is the Gödel number of the quadruple $q_i B|q_i$ and $v(i)$ is the Gödel number of the quadruple $q_i|Rq_{i+1}$. Then take $g(z, y_1, \ldots, y_m)$ to be:

$$[2^{2^9 3^{11} 5^3 7^9} 3^{2^9 3^7 5^3 7^{13}} 5^{2^{13} 3^{11} 5^3 7^9} 7^{2^{13} 3^7 5^7 7^{17}}]*$$

$$\prod_{i=2}^{y_1+2} p_{|2i-4|}^{u(i)} p_{|2i-3|}^{v(i)} * 2^{2^{9+4(y_1+3)} 3^7 5^3 7^{9+4(y_1+4)}} *$$

$$\prod_{i=y_1+4}^{y_1+y_2+4} p_{2|i-(y_1+4)|}^{u(i)} p_{2|i-(y_1+4)|+1}^{v(i)} *$$

$$2^{2^{9+4(y_1+y_2+5)} 3^7 5^3 7^{9+4(y_1+y_2+6)}} * \ldots *$$

$$2^{2^{9+4(y_1+\ldots+y_{m-1}+2m-1)} 3^7 5^3 7^{9+4(y_1+\ldots+y_{m-1}+2m)}} *$$

$$\prod_{i=y_1+\ldots+y_{m-1}+2m}^{y_1+\ldots+y_m+2m} p_{2|i-(y_1+\ldots+y_{m-1}+2m)|}^{u(i)} p_{2|i-(y_1+\ldots+y_{m-1}+2m)|+1}^{v(i)} *$$

$$2^{2^{9+4t} 3^{11} 5^5 7^{9+4t}} 3^{2^{9+4t} 3^7 5^5 7^{9+4(t+1)}} 5^{2^{9+4(t+1)} 3^{11} 5^5 7^{9+4t}} \cdot$$

$$7^{2^{9+4(t+1)} 3^7 5^3 7^{9+4(t+2)}} 11^{2^{9+4(t+2)} 3^7 5^3 7^{9+4(t+3)}} *$$

$$\prod_{i=0}^{\delta(\ell h(z))} p_i^{2^{((z)_i)_0+4(t+3)} 3^{((z)_i)_1} 5^{((z)_i)_2} 7^{((z)_i)_3+4(t+3)}}$$

g is primitive recursive by the results of Section 3.3. When z is not a Gödel number of a Turing machine, φ_z^{m+n} is the empty function and, therefore, $s_n^m(z, y_1, \ldots, y_m)$ must be an index of the empty function and can be taken to be 0. Thus, we define:

$$s_n^m(z, y_1, \ldots, y_m) = \begin{cases} g(z, y_1 \ldots . y_m) & \text{if } TM(z) \\ 0 & \text{otherwise} \end{cases}$$

Hence, s_n^m is primitive recursive.

COROLLARY 5.16

For any partial recursive function $f(x_1, \ldots, x_n, y_1, \ldots, y_m)$, there is a recursive function $g(y_1, \ldots, y_m)$ such that

$$f(x_1, \ldots, x_n, y_1, \ldots, y_m) = \varphi_{g(y_1, \ldots, y_m)}^n (x_1, \ldots, x_n)$$

Proof

Let e be an index of f. By Proposition 5.15,

$$\varphi_e^{m+n}(x_1, \ldots, x_n, y_1, \ldots, y_m) = \varphi_{s_n^m(e, y_1, \ldots, y_m)}(x_1, \ldots, x_n)$$

Let $g(y_1, \ldots, y_m) = s_n^m(e, y_1, \ldots, y_m)$.

Examples

1. Let $G(x)$ be a fixed partial recursive function with non-empty domain. Consider the following decision problem: for any u, determine whether $\varphi_u^1 = G$. Let us show that this problem is recursively unsolvable, that is, that the property $R(u)$, defined by $\varphi_u^1 = G$, is not recursive. Assume, for the sake of contradiction, that R is recursive. Consider the function $f(x,u) = G(x) \cdot N(Z(\mu y T_1(u,u,y)))$. (Recall that $N(Z(t)) = 1$ for all t). Applying Corollary 5.16 to $f(x,u)$, we obtain a recursive function $g(u)$ such that $f(x,u) = \varphi_{g(u)}^1(x)$. For any fixed u, $\varphi_{g(u)}^1 = G$ if and only if $(\exists y)T_1(u,u,y)$. (Here, we use the fact that G has non-empty domain.) Hence, $(\exists y)T_1(u,u,y)$ if and only if $R(g(u))$. Since $R(g(u))$ is recursive, $(\exists y)T_1(u,u,y)$ would be recursive, contradicting Corollary 5.13(a).
2. *A universal Turing machine.* Let the partial recursive function $U(\mu y T_1(z,x,y))$ be computed by a Turing machine $\mathscr{V}$ with Gödel number e. Thus, $U(\mu y T_1(z,x,y)) = U(\mu y T_2(e,z,x,y))$. $\mathscr{V}$ is *universal* in the following sense. First, it can compute every partial recursive function $f(x)$ of one variable. If z is a Gödel number of a Turing machine that computes f, then, if $\mathscr{V}$ begins on the tape $\overline{(z,x)}$, it will compute $U(\mu y T_1(z,x,y)) = f(x)$. Further, $\mathscr{V}$ can be used to compute *any* partial recursive function $h(x_1,\ldots,x_n)$. Let v_0 be a Gödel number of a Turing machine that computes h, and let $f(x) = h((x)_0,(x)_1,\ldots,(x)_{n-1})$. Then $h(x_1,\ldots,x_n) = f(p_0^{x_1}\ldots p_{n-1}^{x_n})$. By applying Corollary 5.16 to the partial recursive function $U(\mu y\, T_n(v,(x)_0,(x)_1,\ldots,(x)_{n-1},y))$, we obtain a recursive function $g(v)$ such that $U(\mu y\, T_n(v,(x)_0,(x)_1,\ldots,(x)_{n-1},y)) = \varphi_{g(v)}^1(x)$. Hence, $f(x) = \varphi_{g(v)}^1(x)$. So $h(x_1,\ldots,x_n)$ is computed by applying $\mathscr{V}$ to the tape $\overline{(g(v_0),p_0^{x_1}\ldots p_{n-1}^{x_n})}$.

Exercises

5.32 Find a *superuniversal* Turing machine $\mathscr{V}^*$ such that, for any Turing machine $\mathscr{T}$, if z is a Gödel number of $\mathscr{T}$ and x is the Gödel number of an initial tape description α of $\mathscr{T}$, then $\mathscr{V}^*$ is applicable to $q_0\overline{(z,x)}$ if and only if $\mathscr{T}$ is applicable to α; moreover, if $\mathscr{T}$, when applied to α, ends with a tape description that has Gödel number w, then $\mathscr{V}^*$, when applied to $q_0\overline{(z,x)}$, produces $\overline{w}$.

5.33 Show that the following decision problem is recursively unsolvable: for any u and v, determine whether $\varphi_u^1 = \varphi_v^1$.

5.34 Show that the following decision problem is recursively unsolvable: for any u, determine whether φ_u^1 has empty domain. (Hence, the condition in Example 1 above, that $G(x)$ has non-empty domain is unnecessary).

5.35

(a) Prove that there is a recursive function $g(u,v)$ such that

$$\varphi_{g(u,v)}^1(x) = \varphi_u^1(x) \cdot \varphi_v^1(x)$$

(b) Prove that there is a recursive function $C(u,v)$ such that

$$\varphi^1_{C(u,v)}(x) = \varphi^1_u(\varphi^1_v(x))$$

5.4 THE KLEENE–MOSTOWSKI HIERARCHY. RECURSIVELY ENUMERABLE SETS

Consider the following array, where $R(x_1, \ldots, x_n, y_1, \ldots, y_m)$ is a recursive relation.

	$R(x_1, \ldots, x_n)$
$(\exists y_1)R(x_1, \ldots, x_n, y_1)$	$(\forall y_1)R(x_1, \ldots, x_n, y_1)$
$(\exists y_1)(\forall y_2)R(x_1, \ldots, x_n, y_1, y_2)$	$(\forall y_1)(\exists y_2)R(x_1, \ldots, x_n, y_1, y_2)$
$(\exists y_1)(\forall y_2)(\exists y_3)R(x_1, \ldots, x_n, y_1, y_2, y_3)$	$(\forall y_1)(\exists y_2)(\forall y_3)R(x_1, \ldots, x_n, y_1, y_2, y_3)$
$\vdots$	$\vdots$

Let $\sum^n_0 = \Pi^n_0 =$ the set of all n-place recursive relations. For $k > 0$, let $\sum^n_k$ be the set of all n-place relations expressible in the prenex form $(\exists y_1)(\forall y_2) \ldots (Qy_k)R(x_1, \ldots, x_n, y_1, \ldots, y_k)$, consisting of k alternating quantifiers beginning with an existential quantifier and followed by a recursive relation R. (Here, '(Qy_k)' denotes $(\exists y_k)$ or $(\forall y_k)$, depending on whether k is odd or even.) Let Π^n_k be the set of all n-place relations expressible in the prenex form $(\forall y_1)(\exists y_2) \ldots (Qy_k)R(x_1, \ldots, x_n, y_1, \ldots, y_k)$, consisting of k alternating quantifiers beginning with a universal quantifier and followed by a recursive relation R. Then the array above can be written

$\sum^n_0$	
$\sum^n_1$	Π^n_1
$\sum^n_2$	Π^n_2
$\sum^n_3$	Π^n_3
$\vdots$	$\vdots$

This array of classes of relations is called the *Kleene–Mostowski hierarchy*, or the *arithmetical hierarchy*.

PROPOSITION 5.17

(a) Every relation expressible in any form listed above is expressible in all the forms in lower rows; that is, for all $j > k$,

$$\sum^n_k \subseteq \sum^n_j \cap \prod^n_j \quad \text{and} \quad \prod^n_k \subseteq \sum^n_j \cap \prod^n_j$$

(b) There is a relation of each form, except $\sum^n_0$, that is not expressible in the other form in the same row and, hence, by part (a), not in any of the rows above; that is, for $k > 0$,

$$\sum_k^n - \prod_k^n \neq \emptyset \quad \text{and} \quad \prod_k^n - \sum_k^n \neq \emptyset$$

(c) Every arithmetical relation is expressible in at least one of these forms.

(d) (Post) For any relation $Q(x_1, \ldots, x_n)$, Q is recursive if and only if both Q and $\neg Q$ are expressible in the form $(\exists y_1)R(x_1, \ldots, x_n, y_1)$, where R is recursive; that is, $\sum_1^n \cap \prod_1^n = \sum_0^n$.

(e) If $Q_1 \in \sum_k^n$ and $Q_2 \in \sum_k^n$, then $Q_1 \vee Q_2$ and $Q_1 \wedge Q_2$ are in $\sum_k^n$. If $Q_1 \in \prod_k^n$ and $Q_2 \in \prod_k^n$, then $Q_1 \vee Q_2$ and $Q_1 \wedge Q_2$ are in $\prod_k^n$.

(f) In contradistinction to part (d), if $k > 0$, then

$$\left(\sum_{k+1}^n \cap \prod_{k+1}^n\right) - \left(\sum_k^n \cup \prod_k^n\right) \neq \emptyset$$

Proof

(a) $(\exists z_1)(\forall y_1) \ldots (\exists z_k)(\forall y_k)R(x_1, \ldots, x_n, z_1, y_1, \ldots, z_k, y_k) \Leftrightarrow$
$(\forall u)(\exists z_1)(\forall y_1) \ldots (\exists z_k)(\forall y_k)(R(x_1, \ldots, x_n, z_1, y_1, \ldots, z_k, y_k) \wedge u = u) \Leftrightarrow$
$(\exists z_1)(\forall y_1) \ldots (\exists z_k)(\forall y_k)(\exists u)(R(x_1, \ldots, x_n, z_1, y_1, \ldots, z_k, y_k) \wedge u = u)$
Hence, any relation expressible in one of the forms in the array is expressible in both forms in any lower row.

(b) Let us consider a typical case, say $\sum_3^n$. Take the relation $(\exists v)(\forall z)(\exists y)T_{n+2}(x_1, x_1, x_2, \ldots, x_n, v, z, y)$, which is in $\sum_3^n$. Assume that this is in $\prod_3^n$, that is, it is expressible in the form $(\forall v)(\exists z)(\forall y)$ $R(x_1, \ldots, x_n, v, z, y)$, where R is recursive. By Exercise 5.19, this relation is equivalent to $(\forall v)(\exists z)(\forall y)\neg T_{n+2}(e, x_1, \ldots, x_n, v, z, y)$ for some e. When $x_1 = e$, this yields a contradiction.

(c) Every wf of the first-order theory S can be put into prenex normal form. Then, it suffices to note that $(\exists u)(\exists v)R(u, v)$ is equivalent to $(\exists z)R((z)_0, (z)_1)$, and $(\forall u)(\forall v)R(u, v)$ is equivalent to $(\forall z)R((z)_0, (z)_1)$. Hence, successive quantifiers of the same kind can be condensed into one such quantifier.

(d) If Q is recursive, so is $\neg Q$, and, if $P(x_1, \ldots, x_n)$ is recursive, then $P(x_1, \ldots, x_n) \Leftrightarrow (\exists y)(P(x_1, \ldots, x_n) \wedge y = y)$. Conversely, assume Q is expressible as $(\exists y)R_1(x_1, \ldots, x_n, y)$ and $\neg Q$ as $(\exists y)R_2(x_1, \ldots, x_n, y)$, where the relations R_1 and R_2 are recursive. Hence, $(\forall x_1) \ldots (\forall x_n)$ $(\exists y)$ $(R_1(x_1, \ldots, x_n, y) \vee R_2(x_1, \ldots, x_n, y))$. So, $\psi(x_1, \ldots, x_n) = \mu y(R_1(x_1, \ldots, x_n, y) \vee R_2(x_1, \ldots, x_n, y))$ is recursive. Then, $Q(x_1, \ldots, x_n) \Leftrightarrow R_1(x_1, \ldots, x_n, \psi(x_1, \ldots, x_n))$ and, therefore, Q is recursive.

(e) Use the following facts. If x is not free in $\mathscr{A}$.

$\vdash (\exists x)(\mathscr{A} \vee \mathscr{B}) \Leftrightarrow (\mathscr{A} \vee (\exists x)\mathscr{B})$, $\vdash (\exists x)(\mathscr{A} \wedge \mathscr{B}) \Leftrightarrow (\mathscr{A} \wedge (\exists x)\mathscr{B})$,
$\vdash (\forall x)(\mathscr{A} \vee \mathscr{B}) \Leftrightarrow (\mathscr{A} \vee (\forall x)\mathscr{B})$, $\vdash (\forall x)(\mathscr{A} \wedge \mathscr{B}) \Leftrightarrow (\mathscr{A} \wedge (\forall x)\mathscr{B})$

(f) We shall suggest a proof in the case $n = 1$; the other cases are then easy consequences. Let $Q(x) \in \sum_k^1 - \prod_k^1$. Define $P(x)$ as $(\exists z)[(x = 2z \wedge Q(z)) \vee (x = 2z + 1 \wedge \neg Q(z))]$. It is easy to prove that $P \notin \sum_k^1 \cup \prod_k^1$ and that $P \in \sum_{k+1}^1$. Observe that $P(x)$ holds if and only if

$$(\exists z)(x = 2z \wedge Q(z)) \vee ((\exists z)_{z<x}(x = 2z + 1) \wedge (\forall z)(x = 2z + 1 \Rightarrow \neg Q(z)))$$

Hence, $P \in \prod_{k+1}^1$ (Rogers, 1959).

Exercises

5.36 For any relation W of n variables, prove that $W \in \sum_k^n$ if and only if $\overline{W} \in \prod_k^n$, where $\overline{W}$ is the complement of W with respect to the set of all n-tuples of natural numbers.

5.37 For each $k > 0$, find a *universal* relation V_k in $\sum_k^{n+1}$; that is, for any relation W of n variables: (a) if $W \in \sum_k^n$, then there exists z_0 such that, for all $x_1, \ldots, x_n$, $W(x_1, \ldots, x_n)$ if and only if $V_k(z_0, x_1, \ldots, x_n)$; and (b) if $W \in \prod_k^n$, there exists v_0 such that, for all $x_1, \ldots, x_n$, $W(x_1, \ldots, x_n)$ if and only if $\neg V_k(v_0, x_1, \ldots, x_n)$. [*Hint*: Use Exercise 5.19.]

The *s-m-n* theorem (Proposition 5.15) enables us to prove the following basic result of recursion theory.

PROPOSITION 5.18 (RECURSION THEOREM)

If $n > 1$ and $f(x_1, \ldots, x_n)$ is a partial recursive function, then there exists a natural number e such that

$$f(x_1, \ldots, x_{n-1}, e) = \varphi_e^{n-1}(x_1, \ldots, x_{n-1})$$

Proof

Let d be an index of $f(x_1, \ldots, x_{n-1}, s_{n-1}^1(x_n, x_n))$. Then

$$f(x_1, \ldots, x_{n-1}, s_{n-1}^1(x_n, x_n)) = \varphi_d^n(x_1, \ldots, x_{n-1}, x_n)$$

By the *s-m-n* theorem, $\varphi_d^n(x_1, \ldots, x_n) = \varphi_{s_{n-1}^1(d, x_n)}^{n-1}(x_1, \ldots, x_{n-1})$. Let $e = s_{n-1}^1(d, d)$. Then:

$$\begin{aligned} f(x_1, \ldots, x_{n-1}, e) &= f(x_1, \ldots, x_{n-1}, s_{n-1}^1(d, d)) = \varphi_d^n(x_1, \ldots, x_{n-1}, d) \\ &= \varphi_{s_{n-1}^1(d,d)}^{n-1}(x_1, \ldots, x_{n-1}) = \varphi_e^{n-1}(x_1, \ldots, x_{n-1}) \end{aligned}$$

COROLLARY 5.19 (FIXED-POINT THEOREM)

If $h(x)$ is recursive, then there exists e such that $\varphi_e^1 = \varphi_{h(e)}^1$.

Proof

Applying the recursion theorem to $f(x, u) = \varphi^1_{h(u)}(x)$, we obtain number e such that $f(x, e) = \varphi^1_e(x)$. But $f(x, e) = \varphi^1_{h(e)}(x)$.

COROLLARY 5.20 (RICE'S THEOREM) (RICE, 1953)

Let $\mathscr{F}$ be a set consisting of at least one, but not all, partial recursive functions of one variable. Then the set $A = \{u | \varphi^1_u \in \mathscr{F}\}$ is not recursive.

Proof

By hypothesis, there exist numbers u_1 and u_2 such that $u_1 \in A$ and $u_2 \notin A$. Now assume that A is recursive. Define

$$h(x) = \begin{cases} u_1 & \text{if } x \notin A \\ u_2 & \text{if } x \in A \end{cases}$$

Clearly, $h(x) \in A$ if and only if $x \notin A$. h is recursive, by Proposition 3.19. By the fixed-point theorem, there is a number e such that $\varphi^1_e = \varphi^1_{h(e)}$. Then we obtain a contradiction as follows:

$$\begin{aligned} e \in A \quad & \text{if and only if} \quad \varphi^1_e \in \mathscr{F} \\ & \text{if and only if} \quad \varphi^1_{h(e)} \in \mathscr{F} \\ & \text{if and only if} \quad h(e) \in A \\ & \text{if and only if} \quad e \notin A \end{aligned}$$

Rice's theorem can be used to show the recursive unsolvability of various decision problems.

Example
Consider the following decision problem: for any u, determine whether φ^1_u has an infinite domain. Let $\mathscr{F}$ be the set of all partial recursive functions of one variable that have infinite domain. By Rice's theorem, $\{u | \varphi^1_u \in \mathscr{F}\}$ is not recursive. Hence, the problem is recursively undecidable.

Notice that Example 1 on page 332 and Exercise 5.34 can be handled in the same way.

Exercises

5.38 Show that the following decision problems are recursively unsolvable.
(a) For any u, determine whether φ^1_u has infinite range.
(b) For any u, determine whether φ^1_u is a constant function.
(c) For any u, determine whether φ^1_u is recursive.

5.39

(a) Show that there is a number e such that the domain of φ_e^1 is $\{e\}$.

(b) Show that there is a number e such that the domain of φ_e^1 is $\omega - \{e\}$.

5.40 This exercise will show the existence of a recursive function that is not primitive recursive.

(a) Let $[\sqrt{x}]$ be the largest integer less than or equal to $\sqrt{x}$. Show that $[\sqrt{x}]$ is defined by the recursion

$$\kappa(0) = 0$$
$$\kappa(x+1) = \kappa(x) + \overline{\text{sg}}|(x+1) - (\kappa(x)+1)^2|$$

Hence, $[\sqrt{x}]$ is primitive recursive.

(b) The function $\text{Quadrem}(x) = x \dot{-} [\sqrt{x}]^2$ is primitive recursive and represents the difference between x and the largest square less than or equal to x.

(c) Let $\rho(x,y) = ((x+y)^2 + y)^2 + x$, $\rho_1(z) = \text{Quadrem}(z)$, and $\rho_2(z) = \text{Quadrem}([\sqrt{z}])$. These functions are primitive recursive. Prove the following:

(i) $\rho_1(\rho(x,y)) = x$ and $\rho_2(\rho(x,y)) = y$.

(ii) ρ is a one–one function from ω^2 into ω.

(iii) $\rho_1(0) = \rho_2(0) = 0$ and

$$\left.\begin{aligned}\rho_1(x+1) &= \rho_1(x) + 1\\ \rho_2(x+1) &= \rho_2(x)\end{aligned}\right\} \quad \text{if} \quad \rho_1(x+1) \neq 0$$

(iv) Let ρ^2 denote ρ, and, for $n \geq 3$, define $\rho^n(x_1, \ldots, x_n) = \rho(\rho^{n-1}(x_1, \ldots, x_{n-1}), x_n)$. Then each ρ^n is primitive recursive. Define $\rho_i^n(x) = \rho_i^{n-1}(\rho_1(x))$ for $1 \leqslant i \leqslant n-1$, and $\rho_n^n(x) = \rho_2(x)$. Then each $\rho_i^n, 1 \leqslant i \leqslant n$, is primitive recursive, and $\rho_i^n(\rho^n(x_1, \ldots, x_n)) = x_i$. Hence, ρ^n is a one–one function of ω^n into ω. The ρ^ns and the ρ_i^ns are obtained from ρ, ρ_1 and ρ_2 by substitution.

(d) The recursion rule (V) (p. 174) can be limited to the form

$$F(x_1, \ldots, x_{n+1}, 0) = x_{n+1} \qquad (n \geq 0)$$
$$F(x_1, \ldots, x_{n+1}, y+1) = G(x_1, \ldots, x_{n+1}, y, F(x_1, \ldots, x_{n+1}, y))$$

[*Hint*: Given

$$f(x_1, \ldots, x_n, 0) = g(x_1, \ldots, x_n)$$
$$f(x_1, \ldots, x_n, y+1) = h(x_1, \ldots, x_n, y, f(x_1, \ldots, x_n, y))$$

define F as above, letting $G(x_1, \ldots, x_{n+1}, y, z) = h(x_1, \ldots, x_n, y, z)$. Then $f(x_1, \ldots, x_n, y) = F(x_1, \ldots, x_n, g(x_1, \ldots, x_n), y)$.]

(e) Taking $x+y, x \cdot y$, and $[\sqrt{x}]$ as additional initial functions, we can limit the recursion rule to the one-parameter form:

$$F(x,0) = G(x)$$
$$F(x,y+1) = H(x,y,F(x,y))$$

[*Hint*: Let $n \geq 2$. Given

$$f(x_1,\ldots,x_n,0) = g(x_1,\ldots,x_n)$$
$$f(x_1,\ldots,x_n,y+1) = h(x_1,\ldots,x_n,y,f(x_1,\ldots,x_n,y))$$

let $F(u,y) = f(\rho_1^n(u),\ldots,\rho_n^n(u),y)$. Define F by a permissible recursion. (Note that $\delta(x), x \dot{-} y, \rho^n$ and ρ_i^n are available.) $f(x_1,\ldots,x_n,y) = F(\rho^n(x_1,\ldots,x_n),y)$.]

(f) Taking $x+y$, $x \cdot y$, and $[\sqrt{x}]$ as additional initial functions, we can use $h(y,F(x,y))$ instead of $H(x,y,F(x,y))$ in part (e).
[*Hint*: Given

$$F(x,0) = G(x)$$
$$F(x,y+1) = H(x,y,F(x,y))$$

let $F_1(x,y) = \rho(x,F(x,y))$. Then $x = \rho_1(F_1(x,y))$ and $F(x,y) = \rho_2(F_1(x,y))$. Define $F_1(x,y)$ by a permissible recursion.]

(g) Taking $x+y$, $x \cdot y$, and $[\sqrt{x}]$ as additional initial functions, we can limit uses of the recursion rule to the form

$$f(x,0) = x$$
$$f(x,y+1) = h(y,f(x,y))$$

Hint: Given

$$F(x,0) = G(x)$$
$$F(x,y+1) = h(y,F(x,y))$$

define f as above. Then $f(x,y) = f(G(x),y)$.

(h) Taking $x+y$, $x \cdot y$, $[\sqrt{x}]$ and $x \dot{-} y$ as additional initial functions, we can limit uses of the recursion rule to those of the form

$$g(0) = 0$$
$$g(y+1) = H(y,g(y))$$

[*Hint*: First note that $|x-y| = (x \dot{-} y) + (y \dot{-} x)$ and that $[\sqrt{x}]$ is definable by a suitable recursion. Now, given

$$f(x,0) = x$$
$$f(x,y+1) = h(y,f(x,y))$$

let $g(x) = f(\rho_2(x),\rho_1(x))$. Then

$$
\begin{aligned}
g(0) &= f(\rho_2(0), \rho_1(0)) = f(0,0) = 0 \\
g(x+1) &= f(\rho_2(x+1), \rho_1(x+1)) \\
&= \begin{cases} \rho_2(x+1) & \text{if } \rho_1(x+1) = 0 \\ h(\rho_1(x+1) \dot{-} 1, f(\rho_2(x+1), \rho_1(x+1) \dot{-} 1)) & \text{if } \rho_1(x+1) \neq 0 \end{cases} \\
&= \begin{cases} \rho_2(x+1) & \text{if } \rho_1(x+1) = 0 \\ h(\rho_1(x), f(\rho_1(x), \rho_2(x))) & \text{if } \rho_1(x+1) \neq 0 \end{cases} \\
&= \begin{cases} \rho_2(x+1) & \text{if } \rho_1(x+1) = 0 \\ h(\rho_1(x), g(x)) & \text{if } \rho_1(x+1) \neq 0 \end{cases} \\
&= \rho_2(x+1) \cdot \overline{\text{sg}}(\rho_1(x+1)) + h(\rho_1(x), g(x)) \cdot \text{sg}(\rho_1(x+1)) \\
&= H(x, g(x))
\end{aligned}
$$

Then $f(x,y) = g(\rho(y,x))$. (Note that sg is obtainable by a recursion of the appropriate form and $\overline{\text{sg}}(x) = 1 \dot{-} x$.)

(i) In part (h), $H(y, g(y))$ can be replaced by $H(g(y))$.
[*Hint*: Given

$$
\begin{aligned}
g(0) &= 0 \\
g(y+1) &= H(y, g(y))
\end{aligned}
$$

let $f(u) = \rho(u, g(u))$ and $\varphi(w) = \rho(\rho_1(w) + 1, H(\rho_1(w), \rho_2(w)))$. Then

$$
\begin{aligned}
f(0) &= 0 \\
f(y+1) &= \varphi(f(y))
\end{aligned}
$$

and $g(u) = \rho_2(f(u))$. (Note that sg(x) is given by a recursion of the specified form.)

(j) Show that the equations

$$
\begin{aligned}
\psi(x,0) &= x+1 \\
\psi(0, y+1) &= \psi(1,y) \\
\psi(x+1, y+1) &= \psi(\psi(x, y+1), y)
\end{aligned}
$$

define a number-theoretic function. In addition, prove:

(I) $\psi(x,y) > x$.

(II) $\psi(x,y)$ is monotonic in x, that is, if $x < z$, then $\psi(x,y) < \psi(z,y)$.

(III) $\psi(x+1, y) \leqslant \psi(x, y+1)$.

(IV) $\psi(x,y)$ is monotonic in y, that is, if $y < z$, then $\psi(x,y) < \psi(x,z)$.

(V)$^{\text{D}}$ Use the recursion theorem to show that ψ is recursive. [*Hint*: Use Exercise 5.21 to show that there is a partial recursive function g such that $g(x,0,u) = x+1, g(0, y+1, u) = \varphi_u^2(1,y)$, and $g(x+1, y+1, u) = \varphi_u^2(\varphi_u^2(x, y+1), y)$. Then use the recursion theorem to find e such that $g(x,y,e) = \varphi_e^2(x,y)$. By induction, show that $g(x,y,e) = \psi(x,y)$.]

(VI) For every primitive recursive function $f(x_1, \ldots, x_n)$, there is some fixed m such that

$$f(x_1, \ldots, x_n) < \psi(\max(x_1, \ldots, x_n), m)$$

for all $x_1, \ldots, x_n$. [*Hint*: Prove this first for the initial functions Z, N, U_i^n, $x+y$, $x \times y$, $[\sqrt{x}]$ and $x \dot{-} y$, and then show that it is preserved by substitution and the recursion of part (i).] Hence, for every primitive recursive function $f(x)$, there is some m such that $f(x) < \psi(x, m)$ for all x.

(VII) Prove that $\psi(x,x) + 1$ is recursive but not primitive recursive.

For other proofs of the existence of recursive functions that are not primitive recursive, see Ackermann (1928), Péter (1935; 1967), and R.M. Robinson (1948).

A set of natural numbers is said to be *recursively enumerable* (r.e.) if and only if it is either empty or the range of a recursive function. If we accept Church's thesis, a non-empty recursively enumerable set is a collection of natural numbers generated by some mechanical process or effective procedure.

PROPOSITION 5.21

(a) A set B is r.e. if and only if $x \in B$ is expressible in the form $(\exists y)R(x,y)$, where R is recursive. (We even can allow R here to be primitive recursive.)

(b) B is r.e. if and only if B is either empty or the range of a partial recursive function.[†]

(c) B is r.e. if and only if B is the domain of a partial recursive function.

(d) B is recursive if and only if B and its complement $\overline{B}$ are r.e.[‡]

(e) The set $K = \{x|(\exists y)T_1(x,x,y)\}$ is r.e. but not recursive.

Proof

(a) Assume B is r.e. If B is empty, then $x \in B \Leftrightarrow (\exists y)\ (x \neq x \wedge y \neq y)$. If B is non-empty, then B is the range of a recursive function g. Then $x \in B \Leftrightarrow (\exists y)(g(y) = x)$. Conversely, assume $x \in B \Leftrightarrow (\exists y)R(x,y)$, where R is recursive. If B is empty, then B is r.e. If B is non-empty, then let k be a fixed element of B. Define

$$\theta(z) = \begin{cases} k & \text{if } \neg R((z)_0, (z)_1) \\ (z)_0 & \text{if } R((z)_0, (z)_1) \end{cases}$$

θ is recursive by Proposition 3.19. Clearly, B is the range of θ. (We can take R to be primitive recursive, since, if R is recursive, then, by Corollary 5.12(a), $(\exists y)R(x,y) \Leftrightarrow (\exists y)T_1(e,x,y)$ for some e, and $T_1(e,x,y)$ is primitive recursive.)

[†]Since the empty function is partial recursive and has the empty set as its range, the condition that B is empty can be omitted.

[‡]$\overline{B} = \omega - B$, where ω is the set of natural numbers.

(b) Assume B is the range of a partial recursive function g. If B is empty, then B is r.e. If B is non-empty, then let k be a fixed element of B. By Corollary 5.11, there is a number e such that $g(x) = U(\mu y T_1(e,x,y))$. Let

$$h(z) = \begin{cases} U((z)_1) & \text{if } T_1(e,(z)_0,(z)_1) \\ k & \text{if } \neg T_1(e,(z)_0,(z)_1) \end{cases}$$

By Proposition 3.19, h is primitive recursive. Clearly, B is the range of h. Hence, B is r.e.

(c) Assume B is r.e. If B is empty, then B is the domain of the partial recursive function $\mu y(x + y + 1 = 0)$. If B is non-empty, then B is the range of a recursive function g. Let G be the partial recursive function such that $G(y) = \mu x(g(x) = y)$. Then B is the domain of G. Conversely, assume B is the domain of a partial recursive function H. Then there is a number e such that $H(x) = U(\mu y T_1(e,x,y))$. Hence, $H(x) = z$ if and only if $(\exists y)(T_1(e,x,y) \wedge U(y) = z)$. But, $x \in B$ if and only if $(\exists z)(H(x) = z)$. So, $x \in B$ if and only if $(\exists z)(\exists y)(T_1(e,x,y) \wedge U(y) = z)$, and the latter is equivalent to $(\exists u)(T_1(e,x,(u)_1) \wedge U((u)_1) = (u)_0)$. Moreover, $T_1(e,x,(u)_1) \wedge U((u)_1) = (u)_0$ is recursive. Thus, by part (a), B is r.e.

(d) Use part (a) and Proposition 5.17(d). (The intuitive meaning of part (d) is the following: if there are mechanical procedures for generating B and $\overline{B}$, then to determine whether any number n is in B we need only wait until n is generated by one of the procedures and then observe which procedure produced it.)

(e) Use parts (a) and (d) and Corollary 5.13(a).

Remember that the functions $\varphi_n^1(x) = U(\mu y T_1(n,x,y))$ form an enumeration of all partial recursive functions of one variable. If we designate the domain of φ_n^1 by W_n, then Proposition 5.21(c) tells us that $W_0, W_1, W_2, \ldots$ is an enumeration (with repetitions) of all r.e. sets. The number n is called the *index* of the set W_n.

Exercises

5.41 Prove that a set B is r.e. if and only if it is either empty or the range of a primitive recursive function. [*Hint*: See the proof of Proposition 5.21(b).]

5.42

(a) Prove that the inverse image of a r.e. set B under a partial recursive function f is r.e. (that is, $\{x | f(x) \in B\}$ is r.e.).

(b) Prove that the inverse image of a recursive set under a recursive function is recursive.

(c) Prove that the image of a r.e. set under a partial recursive function is r.e.

(d) Using Church's thesis, give intuitive arguments for the results in parts (a)–(c).

(e) Show that the image of a recursive set under a recursive function need not be recursive.

5.43 Prove that an infinite set is recursive if and only if it is the range of a strictly increasing recursive function. (g is *strictly increasing* if $x < y$ implies $g(x) < g(y)$.)

5.44 Prove that an infinite set is r.e. if and only if it is the range of a one–one recursive function.

5.45 Prove that every infinite r.e. set contains an infinite recursive subset.

5.46 Assume that A and B are r.e. sets.

(a) Prove that $A \cup B$ is r.e. [In fact, show that there is a recursive function $g(u, v)$ such that $W_{g(u,v)} = W_u \cup W_v$.]

(b) Prove that $A \cap B$ is r.e. [In fact, show that there is a recursive function $h(u, v)$ such that $W_{h(u,v)} = W_u \cap W_v$.]

(c) Show that $\overline{A}$ need not be r.e.

(d) Prove that $\bigcup_{n \in A} W_n$ is r.e.

5.47 Show that the assertion

($\bigtriangledown$) A set B is r.e. if and only if B is effectively enumerable (that is, there is a mechanical procedure for generating the numbers in B)

is equivalent to Church's thesis.

5.48 Prove that the set $A = \{u | W_u = \omega\}$ is not r.e.

5.49 A set B is called *creative* if and only if B is r.e. and there is a partial recursive function h such that, for any n, if $W_n \subseteq \overline{B}$, then $h(n) \in \overline{B} - W_n$.

(a) Prove that $\{x | (\exists y) T_1(x, x, y)\}$ is creative.

(b) Show that every creative set is non-recursive.

5.50^{D} A set B is called *simple* if B is r.e., $\overline{B}$ is infinite, and $\overline{B}$ contains no infinite r.e. set. Clearly, every simple set is non-recursive. Show that a simple set exists.

5.51 A *recursive permutation* is a one–one recursive function from ω onto ω. Sets A and B are called *isomorphic* (written $A \simeq B$) if there is a recursive permutation that maps A onto B.

(a) Prove that the recursive permutations form a group under the operation of composition.

(b) Prove that $\simeq$ is an equivalence relation.

(c) Prove that, if A is recursive (r.e., creative, simple) and $A \simeq B$, then B is recursive (r.e., creative, simple).

Myhill (1955) proved that any two creative sets are isomorphic. (See also Bernays, 1957.)

5.52 A is *many–one reducible* to B (written $A \mathrm{R}_{\mathrm{m}} B$) if there is a recursive function f such that $u \in A$ if and only if $f(u) \in B$. (Many–one reducibility of A to B implies that, if the decision problem for membership in B is recursively solvable, so is the decision problem for membership in A.) A and B are

called *many–one equivalent* (written $A \equiv_m B$) if $A R_m B$ and $B R_m A$. A is *one–one reducible* to B (writtten $A R_1 B$) if there is a one–one recursive function f such that $u \in A$ if and only if $f(u) \in B$. A and B are called *one–one equivalent* (written $A \equiv_1 B$) if $A R_1 B$ and $B R_1 A$.

(a) Prove that $\equiv_m$ and $\equiv_1$ are equivalence relations.

(b) Prove that, if A is creative, B is r.e., and $A R_m B$, then B is creative. [Myhill (1955) showed that, if A is creative and B is r.e., then $B R_m A$.]

(c) (Myhill, 1955) Prove that, if $A R_1 B$ then $A R_m B$, and if $A \equiv_1 B$ then $A \equiv_m B$. However, many–one reducibility does not imply one–one reducibility, and many–one equivalence does not imply one–one equivalence. [*Hint*: Let A be a simple set, C an infinite recursive subset of A, and $B = A - C$. Then $A R_1 B$ and $B R_m A$ but not-$(B R_1 A)$.] It can be shown that $A \equiv_1 B$ if and only if $A \simeq B$.

5.53 (Dekker, 1955) A is said to be *productive* if there is a partial recursive function f such that, if $W_n \subseteq A$, then $f(n) \in A - W_n$. Prove the following.

(a) If A is productive, then A is not r.e.; hence, both A and $\overline{A}$ are infinite.

(b)[D] If A is productive, then A has an infinite r.e. subset. Hence, if A is productive, $\overline{A}$ is not simple.

(c) If A is r.e., then A is creative if and only if $\overline{A}$ is productive.

(d)[D] There exist $2^{\aleph_0}$ productive sets.

5.54 (Dekker and Myhill, 1960) A is *recursively equivalent* to B (written $A \sim B$) if there is a one–one partial recursive function that maps A onto B.

(a) Prove that $\sim$ is an equivalence relation.

(b) A is said to be *immune* if A is infinite and A has no infinite r.e. subset. A is said to be *isolated* if A is not recursively equivalent to a proper subset of A. (The isolated sets may be considered the counterparts of the Dedekind-finite sets.) Prove that an infinite set is isolated if and only if it is immune.

(c)[D] Prove that there exist $2^{\aleph_0}$ immune sets.

Recursively enumerable sets play an important role in logic because, if we assume Church's thesis, the set T_K of Gödel numbers of the theorems of any axiomatizable first-order theory K is r.e. (The same holds true of arbitrary formal axiomatic systems.) In fact, the relation (see page 198)

$\mathrm{Pf}_K(y,x)$: y is the Gödel number of a proof in K of a wf with Gödel number x

is recursive if the set of Gödel numbers of the axioms is recursive, that is, if there is a decision procedure for axiomhood and Church's thesis holds. Now, $x \in T_K$ if and only if $(\exists y)\mathrm{Pf}_K(y,x)$ and, therefore, T_K is r.e. Thus, if we accept Church's thesis, K is decidable if and only if the r.e. set T_K is recursive. It was shown in Corollary 3.46 that every consistent extension K of the theory RR is recursively undecidable, that is, T_K is not recursive.

Much more general results along these lines can be proved (see Smullyan, 1961; Feferman, 1957; Putnam, 1957; Ehrenfeucht and Feferman, 1960; and Myhill, 1955). For example, if K is a first-order theory with equality in the language $\mathscr{L}_A$ of arithmetic: (1) if every recursive set is expressible in K, then K is essentially recursively undecidable, that is, for every consistent extension K′ of K, $T_{K'}$ is not recursive (see Exercise 5.58); (2) if every recursive function is representable in K and K satisfies conditions 4 and 5 on page 208, then the set T_K is creative. For further study of r.e. sets, see Post (1944) and Rogers (1967); for the relationship between logic and recursion theory, see Yasuhara (1971) and Monk (1976, part III).

Exercises

5.55 Let K be a first-order theory with equality in the language $\mathscr{L}_A$ of arithmetic. A number-theoretic relation $B(x_1, \ldots, x_n)$ is said to be *weakly expressible* in K if there is a wf $\mathscr{B}(x_1, \ldots, x_n)$ of K such that, for any natural numbers $k_1, \ldots, k_n, B(k_1, \ldots, k_n)$ if and only if $\vdash_K \mathscr{B}(\bar{k}_1, \ldots, \bar{k}_n)$.

(a) Show that, if K is consistent, then every relation expressible in K is weakly expressible in K.

(b) Prove that, if every recursive relation is expressible in K and K is ω-consistent, every r.e. set is weakly expressible in K. (Recall that, when we refer here to a r.e. set B, we mean the corresponding relation '$x \in B$'.)

(c) If K has a recursive vocabulary and a recursive axiom set, prove that any set that is weakly expressible in K is r.e.

(d) If formal number theory S is ω-consistent, prove that a set B is r.e. if and only if B is weakly expressible in S.

5.56

(a) (Craig, 1953) Let K be a first-order theory such that the set T_K of Gödel numbers of theorems of K is r.e. Show that K is recursively axiomatizable.

(b) For any wf $\mathscr{B}$ of formal number theory S, let $\mathscr{B}\#$ represent its translation into axiomatic set theory NBG (see page 269. Prove that the set of wfs $\mathscr{B}$ such that $\vdash_{NBG} \mathscr{B}\#$ is a (proper) recursively axiomatizable extension of S. (However, no 'natural' set of axioms for this theory is known.)

5.57 Given a set A of natural numbers, let $u \in A^*$ if and only if u is a Gödel number of a wf $\mathscr{B}(x_1)$ and the Gödel number of $\mathscr{B}(\bar{u})$ is in A. Prove that, if A is recursive, then A^* is recursive.

5.58 Let K be a consistent theory in the language $\mathscr{L}_A$ of arithmetic.

(a) Prove that $(\overline{T}_K)^*$ is not weakly expressible in K.

(b) If every recursive set is weakly expressible in K, show that K is recursively undecidable.

(c) If every recursive set is expressible in K, prove that K is essentially recursively undecidable.

5.5 OTHER NOTIONS OF COMPUTABILITY

Computability has been treated here in terms of Turing machines because Turing's definition is probably the one that makes clearest the equivalence between the precise mathematical concept and the intuitive notion.† We already have encountered other equivalent notions: standard Turing computability and partial recursiveness. One of the strongest arguments for the rightness of Turing's definition is that all of the many definitions that have been proposed have turned out to be equivalent. We shall present several of these other definitions.

Herbrand–Gödel Computability

The idea of defining computable functions in terms of fairly simple systems of equations was proposed by Herbrand, given a more precise form by Gödel (1934), and developed in detail by Kleene (1936a). The exposition given here is a version of the presentation in Kleene (1952, chap. XI.)

First let us define the *terms*.

1. All variables are terms.
2. 0 is a term.
3. If t is a term, then $(t)'$ is a term.
4. If $t_1, \ldots, t_n$ are terms and f_j^n is a function letter, then $f_j^n(t_1, \ldots, t_n)$ is a term.

For every natural number n, we define the corresponding *numeral* $\overline{n}$ as follows: (1) $\overline{0}$ is 0 and (2) $\overline{n+1}$ is $(\overline{n})'$. Thus, every numeral is a term.

An *equation* is a formula $r = s$ where r and s are terms. A *system* E of equations is a finite sequence $r_1 = s_1, r_2 = s_2, \ldots, r_k = s_k$ of equations such that r_k is of the form $f_j^n(t_1, \ldots, t_n)$.

The function letter f_j^n is called the *principal letter* of the system E. Those function letters (if any) that appear only on the right-hand side of equations of E are called the *initial letters* of E; any function letter other than the principal letter that appears on the left-hand side of some equations and also on the right-hand side of some equations is called an *auxiliary letter* of E.

We have two rules of inference:

†For further justification of this equivalence, see Turing (1936–37), Kleene (1952, pp. 317–323, 376–381) and Mendelson (1990).

R_1: An equation e_2 is a consequence of an equation e_1 by R_1 if and only if e_2 arises from e_1 by substituting any numeral $\overline{n}$ for all occurrences of a variable.

R_2: An equation e is a consequence by R_2 of equations $f_h^m(\overline{n}_1, \ldots, \overline{n}_m) = \overline{p}$ and $r = s$ if and only if e arises from $r = s$ by replacing one or more occurrences of $f_h^m(\overline{n}_1, \ldots, \overline{n}_m)$ in s by $\overline{p}$, and $r = s$ contains no variables.

A *proof* of an equation e from a set B of equations is a sequence $e_0, \ldots, e_n$ of equations such that e_n is e and, if $0 \leqslant i \leqslant n$, then: (1) e_i is an equation of B, or (2) e_i is a consequence by R_1 of a preceding equation $e_j (j < i)$, or (3) e_i is a consequence by R_2 of two preceding equations e_j and $e_m (j < i, m < i)$. We use the notation $B \vdash e$ to state that there is a proof from B of e (or, in other words, that e is *derivable* from B).

Example

Let E be the system

$$f_1^1(x_1) = (x_1)'$$
$$f_1^2(x_1, x_2) = f_1^3(\overline{2}, x_2, f_1^1(x_1))$$

The principal letter of E is f_1^2, f_1^1 is an auxiliary letter, and f_1^3 is an initial letter. The sequence of equations

$$f_1^2(x_1, x_2) = f_1^3(\overline{2}, x_2, f_1^1(x_1))$$
$$f_1^2(\overline{2}, x_2) = f_1^3(\overline{2}, x_2, f_1^1(\overline{2}))$$
$$f_1^2(\overline{2}, \overline{1}) = f_1^3(\overline{2}, \overline{1}, f_1^1(\overline{2}))$$
$$f_1^1(x_1) = (x_1)'$$
$$f_1^1(\overline{2}) = (\overline{2})' \qquad \text{(i.e., } f_1^1(\overline{2}) = \overline{3})$$
$$f_1^2(\overline{2}, \overline{1}) = f_1^3(\overline{2}, \overline{1}, \overline{3})$$

is a proof of $f_1^2(\overline{2}, \overline{1}) = f_1^3(\overline{2}, \overline{1}, \overline{3})$ from E.

A number-theoretic partial function $\varphi(x_1, \ldots, x_n)$ is said to be *computed* by a system E of equations if and only if the principal letter of E is a letter f_j^n and, for any natural numbers $k_1, \ldots, k_n, p$,

$$E \vdash f_j^n(\overline{k}_1, \ldots, \overline{k}_n) = \overline{p} \text{ if and only if } \varphi(k_1, \ldots, k_n) = p$$

The function φ is called *Herbrand–Gödel-computable* (for short, HG-computable) if and only if there is a system E of equations by which φ is computed.

Examples

1. Let E be the system $f_1^1(x_1) = 0$. Then E computes the zero function Z. Hence, Z is HG-computable.
2. Let E be the system $f_1^1(x_1) = (x_1)'$. Then E computes the successor function N. Hence, N is HG-computable.

3. Let E be the system $f_i^n(x_1, \ldots, x_n) = x_i$. Then E computes the projection function U_i^n. Hence, U_i^n is HG-computable.
4. Let E be the system

$$f_1^2(x_1, 0) = x_1$$
$$f_1^2(x_1, (x_2)') = (f_1^2(x_1, x_2))'$$

Then E computes the addition function.

5. Let E be the system

$$f_1^1(x_1) = 0$$
$$f_1^1(x_1) = x_1$$

The function $\varphi(x_1)$ computed by E is the partial function with domain $\{0\}$ such that $\varphi(0) = 0$. For every $k \neq 0$, $E \vdash f_1^1(\bar{k}) = \bar{0}$ and $E \vdash f_1^1(\bar{k}) = \bar{k}$. Hence, $\varphi(x_1)$ is not defined for $x_1 \neq 0$.

Exercises

5.59

(a) What functions are HG-computable by the following systems of equations ?

(i) $f_1^1(0) = 0, \quad f_1^1((x_1)') = x_1$
(ii) $f_1^2(x_1, 0) = x_1, \quad f_1^2(0, x_2) = 0, \quad f_1^2((x_1)', (x_2)') = f_1^2(x_1, x_2)$
(iii) $f_1^1(x_1) = 0, \quad f_1^1(x_1) = 0'$
(iv) $f_1^2(x_1, 0) = x_1, \quad f_1^2(x_1, (x_2)') = (f_1^2(x_1, x_2))', \quad f_1^1(x_1) = f_1^2(x_1, x_1)$

(b) Show that the following functions are HG-computable.

(i) $|x_1 - x_2|$
(ii) $x_1 \cdot x_2$
(iii) $\varphi(x) = \begin{cases} 0 & \text{if } x \text{ is even} \\ 1 & \text{if } x \text{ is odd} \end{cases}$

5.60

(a) Find a system E of equations that computes the n-place function that is nowhere defined.
(b) Let f be an n-place function defined on a finite domain. Find a system of equations that computes f.
(c) If $f(x)$ is an HG-computable total function and $g(x)$ is a partial function that coincides with $f(x)$ except on a finite set A, where g is undefined, find a system of equations that computes g.

PROPOSITION 5.22

Every partial recursive function is HG-computable.

Proof

(a) Examples 1–3 above show that the initial functions Z, N and U_i^n are HG-computable.

(b) (*Substitution rule* (IV).) Let $\varphi(x_1,\ldots,x_n) = \eta(\psi_1(x_1,\ldots,x_n),\ldots,\psi_m(x_1,\ldots,x_n))$ where η, $\psi_1,\ldots,\psi_m$ have been shown to be HG-computable. Let E_i be a system of equations computing ψ_i, with principal letter f_i^n, and let E_{m+1} be a system of equations computing η, with principal letter f_{m+1}. By changing indices we may assume that no two of $E_1,\ldots,E_{m+1}$ have any function letters in common. Construct a system E for φ by listing $E_1,\ldots,E_{m+1}$ and then adding the equation $f_{m+2}^n(x_1,\ldots,x_n) = f_{m+1}^m(f_1^n(x_1,\ldots,x_n),\ldots,f_m^n(x_1,\ldots,x_n))$. (We may assume that f_{m+2}^n does not occur in $E_1,\ldots,E_{m+1}$.) It is clear that, if $\varphi(k_1,\ldots,k_n) = p$, then $E \vdash f_{m+2}^n(\bar{k}_1,\ldots,\bar{k}_n) = \bar{p}$. Conversely, if $E \vdash f_{m+2}^n(\bar{k}_1,\ldots,\bar{k}_n) = \bar{p}$, then $E \vdash f_1^n(\bar{k}_1,\ldots,\bar{k}_n) = \bar{p}_1,\ldots,$ $E \vdash f_m^n(\bar{k}_1,\ldots,\bar{k}_n) = \bar{p}_m$ and $E \vdash f_{m+1}^m(\bar{p}_1,\ldots,\bar{p}_m) = \bar{p}$. Hence, it readily follows that $E_1 \vdash f_1^n(\bar{k}_1,\ldots,\bar{k}_n) = \bar{p}_1,\ldots,$ $E_m \vdash f_m^n(\bar{k}_1,\ldots,\bar{k}_n) = \bar{p}_m$ and $E_{m+1} \vdash f_{m+1}^n(\bar{p}_1,\ldots,\bar{p}_m) = \bar{p}$. Consequently, $\psi_1(k_1,\ldots,k_n) = p_1,\ldots,\psi_m(k_1,\ldots,k_n) = p_m$ and $\eta(p_1,\ldots,p_m) = p$. So, $\varphi(k_1,\ldots,k_n) = p$. [Hints as to the details of the proof may be found in Kleene (1952, chap. XI, especially, pp. 262–270).] Hence, φ is HG-computable.

(c) (*Recursion rule* (V).) Let

$$\varphi(x_1,\ldots,x_n,0) = \psi(x_1,\ldots,x_n)$$
$$\varphi(x_1,\ldots,x_n,x_{n+1}+1) = \vartheta(x_1,\ldots,x_{n+1},\varphi(x_1,\ldots,x_{n+1}))$$

where ψ and ϑ are HG-computable. Assume that E_1 is a system of equations computing ψ with principal letter f_1^n and that E_2 is a system of equations computing ϑ with principal letter f_1^{n+2}. Then form a system for computing φ by adding to E_1 and E_2

$$f_1^{n+1}(x_1,\ldots,x_n,0) = f_1^n(x_1,\ldots,x_n)$$
$$f_1^{n+1}(x_1,\ldots,x_n,(x_{n+1})') = f_1^{n+2}(x_1,\ldots,x_{n+1},f_1^{n+1}(x_1,\ldots,x_{n+1}))$$

(We assume that E_1 and E_2 have no function letters in common.) Clearly, if $\varphi(k_1,\ldots,k_n,k) = p$, then $E \vdash f_1^{n+1}(\bar{k}_1,\ldots,\bar{k}_n,\bar{k}) = \bar{p}$. Conversely, one can prove easily by induction on k that, if $E \vdash f_1^{n+1}(\bar{k}_1,\ldots,\bar{k}_n,\bar{k}) = \bar{p}$, then $\varphi(k_1,\ldots,k_n,k) = p$. Therefore, φ is HG-computable. (The case when the recursion has no parameters is even easier to handle.)

(d) (μ-operator rule (VI).) Let $\varphi(x_1,\ldots,x_n) = \mu y(\psi(x_1,\ldots,x_n,y) = 0)$ and assume that ψ is HG-computable by a system E_1 of equations with principal letter f_1^{n+1}. By parts (a)–(c), we know that every primitive recursive function is HG-computable. In particular, multiplication is HG-computable; hence, there is a system E_2 of equations having no function letters in common with E_1 and with principal letter f_2^2 such that $E_2 \vdash f_2^2(\bar{k}_1,\bar{k}_2) = \bar{p}$ if and only if $k_1 \cdot k_2 = p$. We form a system E_3 by adding to E_1 and E_2 the equations

$$f_3^{n+1}(x_1,\ldots,x_n,0) = 1$$
$$f_3^{n+1}(x_1,\ldots,x_n,(x_{n+1})') = f_2^2(f_3^{n+1}(x_1,\ldots,x_n,x_{n+1}),f_1^{n+1}(x_1,\ldots,x_nx_{n+1}))$$

One can prove by induction that E_3 computes the function $\prod_{y<z}\psi(x_1,\ldots,x_n,y)$; that is, $E_3 \vdash f_3^{n+1}(\bar{k}_1,\ldots,\bar{k}_n,\bar{k}) = \bar{p}$ if and only if $\prod_{y<z}\psi(k_1,\ldots,k_n,y) = p$. Now construct the system E by adding to E_3 the equations

$$f_4^3((x_1)',0,x_3) = x_3$$
$$f_3^n(x_1,\ldots,x_n) = f_4^3(f_3^{n+1}(x_1,\ldots,x_n,x_{n+1}),f_3^{n+1}(x_1,\ldots,x_n,(x_{n+1})'),x_{n+1})$$

Then E computes the function $\varphi(x_1,\ldots,x_n) = \mu y(\psi(x_1,\ldots,x_n,y) = 0)$. If $\mu y(\psi(k_1,\ldots,k_n,y) = 0) = q$, then $E_3 \vdash f_3^{n+1}(\bar{k}_1,\ldots,\bar{k}_n,\bar{q}) = \bar{p}'$, where $p+1 = \prod_{y<q}\psi(k_1,\ldots,k_n,y)$, and $E_3 \vdash f_3^{n+1}(\bar{k}_1,\ldots,\bar{k}_n,\bar{q}') = 0$. Hence, $E \vdash f_3^n(\bar{k}_1,\ldots,\bar{k}_n) = f_4^3(\bar{p}',0,\bar{q})$. But, $E \vdash f_4^3(\bar{p}',0,\bar{q}) = \bar{q}$, and so, $E \vdash f_3^n(\bar{k}_1,\ldots,\bar{k}_n) = \bar{q}$. Conversely, if $E \vdash f_3^n(\bar{k}_1,\ldots,\bar{k}_n) = \bar{q}$, then $E \vdash f_4^3(\bar{m}',0,\bar{q}) = \bar{q}$, where $E_3 \vdash f_3^{n+1}(\bar{k}_1,\ldots,\bar{k}_n,\bar{q}) = (\bar{m})'$ and $E_3 \vdash f_3^{n+1}(\bar{k}_1,\ldots,\bar{k}_n,\bar{q}') = 0$. Hence, $\prod_{y<q}\psi(k_1,\ldots,k_n,y) = m+1 \neq 0$ and $\prod_{y<q+1}\psi(k_1,\ldots,k_n,y) = 0$. So, $\psi(k_1,\ldots,k_n,y) \neq 0$ for $y < q$, and $\psi(k_1,\ldots,k_n,q) = 0$. Thus, $\mu y(\psi(k_1,\ldots,k_n,y) = 0) = q$. Therefore, φ is HG-computable.

We now shall proceed to show that every HG-computable function is partial recursive by means of an arithmetization of the apparatus of Herbrand–Gödel computability. We shall use the same arithmetization that was used for first-order theories (see Section 3.4). (We take the symbol $'$ to be an abbreviation for f_1^1. Remember that $r = s$ is an abbreviation for $A_1^2(r,s)$. The only individual constant is 0.) In particular, the following relations and functions are primitive recursive (see pages 192–4):

FL(x): x is the Gödel number of a function letter

$$(\exists y)_{y<x}(\exists z)_{z<x}(x = 1 + 8(2^y \cdot 3^z) \wedge y > 0 \wedge z > 0)$$

EVbl(x): x is the Gödel number of an expression consisting of a variable

EFL(x): x is the Gödel number of an expression consisting of a function letter

Nu(x): x is the Gödel number of a numeral

Trm(x): x is the Gödel number of a term

Num(x) = the Gödel number of the numeral $\bar{x}$

$\text{Arg}_T(x)$ = the number of arguments of a function letter, f, if x is the Gödel number of f

$x * y$ = the Gödel number of an expression AB if x is the Gödel number of the expression A and y is the Gödel number of B

Subst(x,y,u,v): v is the Gödel number of a variable x_i, u is the Gödel number of a term t, y is the Gödel number of an expression $\mathscr{B}$, and x is the Gödel number of the result of substituting t for all occurrences of x_i in $\mathscr{B}$

The following are also primitive recursive:

Eqt(x): x is the Gödel number of an equation:

$$\ell h(x) = 3 \wedge \mathrm{Trm}((x)_1) \wedge \mathrm{Trm}((x)_2) \wedge (x)_0 = 99$$

(Remember that= is A_1^2, whose Gödel number is 99.)

Syst(x): x is the Gödel number of a system of equations:

$$(\forall y)_{y<\ell h(x)}\mathrm{Eqt}((x)_y) \wedge \mathrm{FL}((((x)_{\ell h(x)\dot{-}1})_1)_0)$$

Occ(u, v): u is the Gödel number of a term t or equation $\mathfrak{B}$ and v is the Gödel number of a term that occurs in t or $\mathfrak{B}$:

$$(\mathrm{Trm}(u) \vee \mathrm{Eqt}(u)) \wedge \mathrm{Trm}(v) \wedge (\exists x)_{x<u}(\exists y)_{y<u}(u = x * v * y \vee u = x * v \vee u = v * y \vee u = v)$$

$\mathrm{Cons}_1(u, v)$: u is the Gödel number of an equation e_1, v is the Gödel number of an equation e_2, and e_2 is a consequence of e_1 by rule R_1:

$$\mathrm{Eqt}(u) \wedge \mathrm{Eqt}(v) \wedge (\exists x)_{x<u}(\exists y)_{y<v}(\mathrm{Nu}(y) \wedge \mathrm{Subst}(v, u, y, x) \wedge \mathrm{Occ}(u, x))$$

$\mathrm{Cons}_2(u, z, v)$: u, z, v are Gödel numbers of equations e_1, e_2, e_3, respectively, and e_3 is a consequence of e_1 and e_2 by rule R_2:

$$\begin{aligned}&\mathrm{Eqt}(u) \wedge \mathrm{Eqt}(z). \wedge \mathrm{Eqt}(v) \wedge \neg(\exists x)_{x<z}(\mathrm{EVbl}(x) \wedge \mathrm{Occ}(z, x))\\ &\wedge \mathrm{FL}(((z)_1)_0) \wedge (\forall x)_{0<x<\ell h((z)_1)}\neg\mathrm{FL}(((z)_1)_x)\\ &\wedge (\forall x)_{x<\ell h((z)_2)}\neg\mathrm{FL}(((z)_2)_x) \wedge \mathrm{Occ}((u)_2, (z)_1)\\ &\wedge [(\exists y)_{y<u}(\exists w)_{w<u}((u)_2 = y * (z)_1 * w \wedge v = 2^{99}3^{(u)_1}5^{y*(z)_2*w}) \vee .\\ &((u)_2 = (z)_1 \wedge v = 2^{99}3^{(u)_1}5^{(z)_2})]\end{aligned}$$

Ded(u, z): u is the Gödel number of a system of equations E and z is the Gödel number of a proof from E:

$$\begin{aligned}&\mathrm{Syst}(u) \wedge (\forall x)_{x<\ell h(z)}((\exists w)_{w<\ell h(u)}(u)_w = (z)_x\\ &\qquad \vee (\exists y)_{y<x}\mathrm{Cons}_1((z)_y, (z)_x) \vee (\exists y)_{y<x}(\exists v)_{v<x}\mathrm{Cons}_2((z)_y, (z)_v, (z)_x))\end{aligned}$$

$\mathrm{S}_n(u, x_1, \ldots, x_n, z)$: u is the Gödel number of a system of equations E whose principal letter is of the form f_j^n, and z is the Gödel number of a proof from E of an equation of the form $f_j^n(\bar{x}_1, \ldots, \bar{x}_n) = \bar{p}$:

$$\begin{aligned}&\mathrm{Ded}(u, z) \wedge \mathrm{Arg}_{\mathrm{T}}(((u)_{\ell h(u)\dot{-}1})_1)_0 = n \wedge (((z)_{\ell h(z)\dot{-}1})_1)_0\\ &= (((u)_{\ell h(u)\dot{-}1})_1)_0 \wedge (\forall y)_{0<y<\ell h(((z)_{\ell h(z)\dot{-}1})_1)}\neg\mathrm{FL}((((z)_{\ell h(z)\dot{-}1})_1)_y)\\ &\wedge \mathrm{Nu}(((z)_{\ell h(z)\dot{-}1})_2) \wedge ((z)_{\ell h(z)\dot{-}1})_1 = 2^{(((u)_{\ell h(u)\dot{-}1})_1)_0} * 2^3 * 2^{\mathrm{Num}(x_1)} * 2^7\\ &* 2^{\mathrm{Num}(x_2)} * 2^7 * \cdots * 2^7 * 2^{\mathrm{Num}(x_n)} * 2^5\end{aligned}$$

Remember that $g(() = 3, g()) = 5$ and $g(,) = 7$.

$U(x) = \mu y_{y<x}(\mathrm{Num}(y) = ((x)_{\ell\ell(x)\dot{-}1})_2)$. (If x is the Gödel number of a proof of an equation $r = \overline{p}$, then $U(x) = p$.)

PROPOSITION 5.23

(Kleene, 1936a) If $\varphi(x_1, \ldots, x_n)$ is HG-computable by a system of equations E with Gödel number e, then

$$\varphi(x_1, \ldots, x_n) = U(\mu y(\mathrm{S}_n(e, x_1, \ldots, x_n, y)))$$

Hence, every HG-computable function φ is partial recursive, and, if φ is total, then φ is recursive.

Proof

$\varphi(k_1, \ldots, k_n) = p$ if and only if $\mathrm{E} \vdash f_j^n(\overline{k}_1, \ldots, \overline{k}_n) = \overline{p}$, where f_j^n is the principal letter of E. $\varphi(k_1, \ldots, k_n)$ is defined if and only if $(\exists y)\mathrm{S}_n(e, k_1, \ldots, k_n, y)$. If $\varphi(k_1, \ldots, k_n)$ is defined, $\mu y(\mathrm{S}_n(e, k_1, \ldots, k_n, y))$ is the Gödel number of a proof from E of an equation $f_j^n(\overline{k}_1, \ldots, \overline{k}_n) = \overline{p}$. Hence, $U(\mu y(\mathrm{S}_n(e, k_1, \ldots, k_n, y))) = p = \varphi(k_1, \ldots, k_n)$. Also, since S_n is primitive recursive, $\mu y(\mathrm{S}_n(e, x_1, \ldots, x_n, y))$ is partial recursive. If φ is total, then $(\forall x_1) \ldots (\forall x_n)(\exists y)\mathrm{S}_n(e, x_1, \ldots, x_n, y)$; hence, $\mu y(\mathrm{S}_n(e, x_1, \ldots, x_n, y))$ is recursive, and then, so is $U(\mu y(\mathrm{S}_n(e, x_1, \ldots, x_n, y)))$.

Thus, the class of HG-computable functions is identical with the class of partial recursive functions. This is further evidence for Church's thesis.

Markov algorithms

By an *algorithm in* an alphabet A we mean a computable function $\mathfrak{A}$ whose domain is a subset of the set of words of A and the values of which are also words in A. If P is a word in A, $\mathfrak{A}$ is said to be *applicable* to P if P is in the domain of $\mathfrak{A}$; if $\mathfrak{A}$ is applicable to P, we denote its value by $\mathfrak{A}(\mathrm{P})$. By an *algorithm over* an alphabet A we mean an algorithm $\mathfrak{A}$ in an extension B of A.† Of course, the notion of algorithm is as hazy as that of computable function.

Most familiar algorithms can be broken down into a few simple steps. Starting from this observation and following Markov (1954), we select a particularly simple operation, substitution of one word for another, as the basic unit from which algorithms are to be constructed. To this end, if P and Q are words of an alphabet A, then we call the expressions $\mathrm{P} \to \mathrm{Q}$ and $\mathrm{P} \to \cdot\, \mathrm{Q}$ *productions* in the alphabet A. We assume that '$\to$' and '$\cdot$' are not symbols of A. Notice that P or Q is permitted to be the empty word. $\mathrm{P} \to \mathrm{Q}$

†An alphabet B is an extension of A if $\mathrm{A} \subseteq \mathrm{B}$.

is called a *simple* production, whereas $P \to \cdot\, Q$ is a *terminal* production. Let us use $P \to (\cdot)\, Q$ to denote either $P \to Q$ or $P \to \cdot\, Q$. A finite list of productions in A

$$\begin{array}{ll} P_1 & \to (\cdot)\ Q_1 \\ P_2 & \to (\cdot)\ Q_2 \\ & \vdots \\ P_r & \to (\cdot)\ Q_r \end{array}$$

is called an *algorithm schema* and determines the following algorithm $\mathfrak{A}$ in A. As a preliminary definition, we say that a word T *occurs* in a word Q if there are words U, V (either one possibly the empty word Λ) such that $Q = UTV$. Now, given a word P in A: (1) We write $\mathfrak{A}: P\sqsupset$ if none of the words $P_1, \ldots, P_r$ occurs in P. (2) Otherwise, if m is the least integer, with $1 \leqslant m \leqslant r$, such that P_m occurs in P, and if R is the word that results from replacing the leftmost occurrence of P_m in P by Q_m, then we write

$$\text{(a)} \quad \mathfrak{A} : P \vdash R$$

if $P_m \to (\cdot)\, Q_m$ is simple (and we say that $\mathfrak{A}$ simply transforms P into R);

$$\text{(b)} \quad \mathfrak{A} : P \vdash \cdot\, R$$

if $P_m \to (\cdot)\, Q_m$ is terminal (and we say that $\mathfrak{A}$ terminally transforms P into R). We then define $\mathfrak{A} : P \models R$ to mean that there is a sequence $R_0, R_1, \ldots, R_k$ such that:

(i) $P = R_0$.
(ii) $R = R_k$.
(iii) For $0 \leqslant j \leqslant k-2$, $\mathfrak{A} : R_j \vdash R_{j+1}$.
(iv) Either $\mathfrak{A} : R_{k-1} \vdash R_k$ or $\mathfrak{A} : R_{k-1} \vdash \cdot\, R_k$. (In the second case, we write $\mathfrak{A} : P \models \cdot\, R$.)

We set $\mathfrak{A}(P) = R$ if and only if either $\mathfrak{A} : P \models \cdot\, R$, or $\mathfrak{A} : P \models R$ and $\mathfrak{A} : R\sqsupset$. The algorithm thus defined is called a *normal* algorithm (or a *Markov* algorithm) in the alphabet A.

The action of $\mathfrak{A}$ can be described as follows: given a word P, we find the first production $P_m \to (\cdot)\, Q_m$ in the schema such that P_m occurs in P. We then substitute Q_m for the leftmost occurrence of P_m in P. Let R_1 be the new word obtained in this way. If $P_m \to (\cdot)\, Q_m$ was a terminal production, the process stops and the value of the algorithm is R_1. If $P_m \to (\cdot)\, Q_m$ was simple, then we apply the same process to R_1 as was just applied to P, and so on. If we ever obtain a word R_i such that $\mathfrak{A} : R_i\sqsupset$, then the process stops and the value $\mathfrak{A}(P)$ is R_i. It is possible that the process just described never stops. In that case, $\mathfrak{A}$ is not applicable to the given word P.

Our exposition of the theory of normal algorithms will be based on Markov (1954).

Examples

1. Let A be the alphabet {b, c}. Consider the schema

$$b \to \cdot \Lambda$$
$$c \to c$$

The normal algorithm $\mathfrak{A}$ defined by this schema transforms any word that contains at least one occurrence of b into the word obtained by erasing the leftmost occurrence of b. $\mathfrak{A}$ transforms the empty word Λ into itself. $\mathfrak{A}$ is not applicable to any non-empty word that does not contain b.

2. Let A be the alphabet $\{a_0, a_1, \ldots, a_n\}$. Consider the schema

$$a_0 \to \Lambda$$
$$a_1 \to \Lambda$$
$$\vdots$$
$$a_n \to \Lambda$$

We can abbreviate this schema as follows:

$$\xi \to \Lambda \quad (\xi \text{ in A})$$

(Whenever we use such abbreviations, the productions intended may be listed in any order.) The corresponding normal algorithm transforms every word into the empty word. For example,
$\mathfrak{A} : a_1a_2a_1a_3a_0 \vdash a_1a_2a_1a_3 \vdash a_2a_1a_3 \vdash a_2a_3 \vdash a_3 \vdash \Lambda$ and $\mathfrak{A} : \Lambda\square$. Hence, $\mathfrak{A}(a_1a_2a_1a_3a_0) = \Lambda$.

3. Let A be an alphabet containing the symbol a_1, which we shall abbreviate |. For natural numbers n, we define $\overline{n}$ inductively as follows: $\overline{0} =|$ and $\overline{n+1} = \overline{n}\,|$. Thus, $\overline{1} = ||$, $\overline{2} =|||$, and so on. The words $\overline{n}$ will be called *numerals*. Now consider the schema $\Lambda \to \cdot\,|$, defining a normal algorithm $\mathfrak{A}$. For any word P in A, $\mathfrak{A}(P) =| P$.† In particular, for every natural number n, $\mathfrak{A}(\overline{n}) = \overline{n+1}$.
4. Let A be an arbitrary alphabet $\{a_0, a_1, \ldots, a_n\}$. Given a word $P = a_{j_0}a_{j_1}\cdots a_{j_k}$, let $\breve{P} = a_{j_k}\cdots a_{j_1}a_{j_0}$ be the *inverse* of P. We seek a normal algorithm $\mathfrak{A}$ such that $\mathfrak{A}(P) = \breve{P}$. Consider the following (abbreviated) algorithm schema in the alphabet $B = A \cup \{\alpha, \beta\}$.

(a) $\alpha\alpha \to \beta$
(b) $\beta\xi \to \xi\beta \quad (\xi \text{ in A})$
(c) $\beta\alpha \to \beta$
(d) $\beta \to \cdot\Lambda$
(e) $\alpha\eta\xi \to \xi\alpha\eta \quad (\xi, \eta \text{ in A})$
(f) $\Lambda \to \alpha$

†To see this, observe that Λ occurs at the beginning of any word P, since $P = \Lambda P$.

This determines a normal algorithm $\mathfrak{A}$ in B. Let $P = a_{j_0}a_{j_1}\cdots a_{j_k}$ be any word in A. Then $\mathfrak{A} : P \vdash \alpha P$ by production (f). Then, $\alpha P \vdash a_{j_1}\alpha a_{j_0}a_{j_2}\ldots a_{j_k} \vdash a_{j_1}a_{j_2}\alpha a_{j_0}a_{j_3}\ldots a_{j_k}\ldots \vdash a_{j_1}a_{j_2}\ldots a_{j_k}\alpha a_{j_0}$, all by production (e). Thus, $\mathfrak{A} : P \models a_{j_1}a_{j_2}\ldots a_{j_k}\alpha a_{j_0}$. Then, by production (f), $\mathfrak{A} : P \models \alpha a_{j_1}a_{j_2}\ldots a_{j_k}\alpha a_{j_0}$. Applying, as before, production (e), $\mathfrak{A} : P \models a_{j_2}a_{j_3}\ldots a_{j_k}\alpha a_{j_1}\alpha a_{j_0}$ Iterating this process, we obtain $\mathfrak{A} : P \models \alpha a_{j_k}\alpha a_{j_{k-1}}\alpha\ldots\alpha a_{j_1}\alpha a_{j_0}$. Then, by production (f), $\mathfrak{A} : P \models \alpha\alpha a_{j_k}\alpha a_{j_{k-1}}\alpha\ldots\alpha a_{j_1}\alpha a_{j_0}$, and, by production (a), $\mathfrak{A} : P \models \beta a_{j_k}\alpha a_{j_{k-1}}\alpha\ldots\alpha a_{j_1}\alpha a_{j_0}$. Applying productions (b) and (c) and finally (d), we arrive at $\mathfrak{A} : P \models \cdot \breve{P}$. Thus, $\mathfrak{A}$ is a normal algorithm over A that inverts every word of A.†

Exercises

5.61 Let A be an alphabet. Describe the action of the normal algorithms given by the following schemas.

(a) Let Q be a fixed word in A and let the algorithm schema be: $\Lambda \to \cdot Q$.

(b) Let Q be a fixed word in A and let α be a symbol not in A. Let $B = A \cup \{\alpha\}$. Consider the schema

$$\begin{aligned} \alpha\xi &\to \xi\alpha \qquad (\xi \text{ in A}) \\ \alpha &\to \cdot Q \\ \Lambda &\to \alpha \end{aligned}$$

(c) Let Q be a fixed word in A. Take the schema

$$\begin{aligned} \xi &\to \Lambda \qquad (\xi \text{ in A}) \\ \Lambda &\to \cdot Q \end{aligned}$$

(d) Let $B = A \cup \{|\}$. Consider the schema

$$\begin{aligned} \xi &\to | \qquad (\xi \text{ in A} - \{|\}) \\ \Lambda &\to \cdot | \end{aligned}$$

5.62 Let A be an alphabet not containing the symbols α, β, γ. Let $B = A \cup \{\alpha\}$ and $C = A \cup \{\alpha, \beta, \gamma\}$.

(a) Construct a normal algorithm $\mathfrak{A}$ in B such that $\mathfrak{A}(\Lambda) = \Lambda$ and $\mathfrak{A}(\xi P) = P$ for any symbol ξ in A and any word P in A. Thus, $\mathfrak{A}$ erases the first letter of any non-empty word in A.

†The distinction between a normal algorithm in A and a normal algorithm over A is important. A normal algorithm in A uses only symbols of A, whereas a normal algorithm over A may use additional symbols not in A. Every normal algorithm in A is a normal algorithm over A, but there are algorithms in A that are determined by normal algorithms over A but that are not normal algorithms in A (for example, the algorithm of Exercise 5.62(d)).

(b) Construct a normal algorithm $\mathfrak{D}$ in B such that $\mathfrak{D}(\Lambda) = \Lambda$ and $\mathfrak{D}(P\xi) = P$ for any symbol ξ in A and any word P in A. Thus, $\mathfrak{D}$ erases the last letter of any non-empty word in A.

(c) Construct a normal algorithm $\mathfrak{C}$ in B such that $\mathfrak{C}(P)$ equals Λ if P contains exactly two occurrences of α and $\mathfrak{C}(P)$ is defined and is not equal to Λ in all other cases.

(d) Construct a normal algorithm $\mathscr{B}$ in C such that, for any word P of A, $\mathfrak{B}(P) = PP$.

5.63 Let A and B be alphabets and let α be a symbol in neither A nor B. For certain symbols $a_1, \ldots, a_k$ in A, let $Q_1, \ldots, Q_k$ be corresponding words in B. Consider the algorithm that associates with each word P of A the word $\mathrm{Sub}_{Q_1 \ldots Q_k}^{a_1 \ldots a_k}(P)$ obtained by simultaneous substitution of each Q_i for a_i $(i = 1, \ldots, k)$. Show that this is given by a normal algorithm in $A \cup B \cup \{\alpha\}$.

5.64 Let $H = \{|\}$ and $M = \{|, B\}$. Every natural number n is represented by its numeral $\overline{n}$, which is a word in H. We represent every k-tuple $(n_1, n_2, \ldots, n_k)$ of natural numbers by the word $\overline{n}_1 B\, \overline{n}_2 B \ldots B\, \overline{n}_k$ in M. We shall denote this word by $\overline{(n_1, n_2, \ldots, n_k)}$. For example, $\overline{(3,1,2)}$ is $||||B||B|||$.

(a) Show that the schema

$$\begin{aligned} B &\to B \\ \alpha\| &\to \alpha| \\ \alpha| &\to \cdot\, | \\ \Lambda &\to \alpha \end{aligned}$$

defines a normal algorithm $\mathfrak{A}_Z$ over M such that $\mathfrak{A}_Z(\overline{n}) = \overline{0}$ for any n, and $\mathfrak{A}_Z$ is applicable only to numerals in M.

(b) Show that the schema

$$\begin{aligned} B &\to B \\ \alpha| &\to \cdot\| \\ \Lambda &\to \alpha \end{aligned}$$

defines a normal algorithm $\mathfrak{A}_N$ over M such that $\mathfrak{A}_N(\overline{n}) = \overline{n+1}$ for all n, and $\mathfrak{A}_N$ is applicable only to numerals in M.

(c) Let $\alpha_1, \ldots, \alpha_{2k}$ be symbols not in M. Let $1 \leqslant j \leqslant k$. Let $\mathscr{S}_j$ be the list

$$\alpha_{2j-1} B \to \alpha_{2j-1} B$$

$$\alpha_{2j-1}| \to \alpha_{2j}|$$

$$\alpha_{2j}| \to \alpha_{2j}$$

$$\alpha_{2j} B \to \alpha_{2j+1}$$

If $1 < j < k$ consider the algorithm schema

$$\begin{array}{c}
\mathscr{S}_1 \\
\vdots \\
\mathscr{S}_{j-1} \\
\alpha_{2j-1}\,\mathrm{B} \to \alpha_{2j-1}\mathrm{B} \\
\alpha_{2j-1}| \to \alpha_{2j}| \\
\alpha_{2j}| \to |\alpha_{2j} \\
\alpha_{2j}\,\mathrm{B} \to \alpha_{2j+1} \\
\mathscr{S}_{j+1} \\
\vdots \\
\mathscr{S}_{k-1} \\
\alpha_{2k-1}\,\mathrm{B} \to \alpha_{2k-1}\mathrm{B} \\
\alpha_{2k-1}| \to \alpha_{2k} \\
\alpha_{2k}| \to \alpha_{2k} \\
\alpha_{2k}\mathrm{B} \to \alpha_{2k}\mathrm{B} \\
\alpha_{2k} \to \cdot\,\Lambda \\
\Lambda \to \alpha_1
\end{array}$$

If $j = 1$, consider the schema

$$\begin{array}{c}
\alpha_1\,\mathrm{B} \to \alpha_1\mathrm{B} \\
\alpha_1| \to \alpha_2| \\
\alpha_2| \to |\alpha_2 \\
\alpha_2\,\mathrm{B} \to \alpha_3 \\
\mathscr{S}_2 \\
\vdots \\
\mathscr{S}_{k-1} \\
\alpha_{2k-1}\,\mathrm{B} \to \alpha_{2k-1}\mathrm{B} \\
\alpha_{2k-1}| \to \alpha_{2k}| \\
\alpha_{2k}\mathrm{B}| \to \alpha_{2k} \\
\alpha_{2k}\mathrm{B} \to \alpha_{2k}\mathrm{B} \\
\alpha_{2k} \to \cdot\,\Lambda \\
\Lambda \to \alpha_1
\end{array}$$

If $j = k$, consider the schema

$$\begin{array}{c}
\mathscr{S}_1 \\
\vdots \\
\mathscr{S}_{k-1} \\
\alpha_{2k-1}\,\mathrm{B} \to \alpha_{2k-1}\mathrm{B} \\
\alpha_{2k-1}| \to \alpha_{2k}| \\
\alpha_{2k}| \to |\alpha_{2k} \\
\alpha_{2k}\,\mathrm{B} \to \alpha_{2k}\mathrm{B} \\
\alpha_{2k} \to \cdot\,\Lambda \\
\Lambda \to \alpha_1
\end{array}$$

Show that the corresponding normal algorithm $\mathfrak{A}_j^k$ is such that $\mathfrak{A}_j^k(\overline{(n_1, \ldots, n_k)}) = \bar{n}_j$; and $\mathfrak{A}_j^k$ is applicable to only words of the form $\overline{(n_1, \ldots, n_k)}$.

(d) Construct a schema for a normal algorithm in M transforming $\overline{(n_1, n_2)}$ into $\overline{|n_1 - n_2|}$.

(e) Construct a normal algorithm in M for addition.

(f) Construct a normal algorithm over M for multiplication.

Given algorithms $\mathfrak{A}$ and $\mathfrak{B}$ and a word P, we write $\mathfrak{A}(\mathrm{P}) \approx \mathfrak{B}(\mathrm{P})$ if and only if either $\mathfrak{A}$ and $\mathfrak{B}$ are both applicable to P and $\mathfrak{A}(\mathrm{P}) = \mathfrak{B}(\mathrm{P})$ or neither $\mathfrak{A}$ nor $\mathfrak{B}$ is applicable to P. More generally, if C and D are expressions, then $\mathrm{C} \approx \mathrm{D}$ is to hold if and only if neither C nor D is defined, or both C and D are defined and denote the same object. If $\mathfrak{A}$ and $\mathfrak{B}$ are algorithms over an alphabet A, then we say that $\mathfrak{A}$ and $\mathfrak{B}$ are *fully equivalent* relative to A if and only if $\mathfrak{A}(\mathrm{P}) \approx \mathfrak{B}(\mathrm{P})$ for every word P in A; we say that $\mathfrak{A}$ and $\mathfrak{B}$ are *equivalent* relative to A if and only if, for any word P in A, whenever $\mathfrak{A}(\mathrm{P})$ or $\mathfrak{B}(\mathrm{P})$ exists and is in A, then $\mathfrak{A}(\mathrm{P}) \approx \mathfrak{B}(\mathrm{P})$.

Let M be the alphabet $\{|, \mathrm{B}\}$, as in Exercise 5.64, and let ω be the set of natural numbers. Given a partial number-theoretic function φ of k arguments, that is, a function from a subset of ω^k into ω, we denote by $\mathfrak{B}_\varphi$ the corresponding function in M; that is, $\mathfrak{B}_\varphi(\overline{(n_1, \ldots, n_k)}) = \overline{\varphi(n_1, \ldots, n_k)}$ whenever either of the two sides of the equation is defined. $\mathfrak{B}_\varphi$ is assumed to be inapplicable to words not of the form $\overline{(n_1, \ldots, n_k)}$. The function φ is said

to be *Markov-computable* if and only if there is a normal algorithm $\mathfrak{A}$ over M that is fully equivalent to $\mathfrak{B}_\varphi$ relative to M.†

A normal algorithm is said to be *closed* if and only if one of the productions in its schema has the form $\Lambda \rightarrow \cdot\, Q$. Such an algorithm can end only terminally – that is, by an application of a terminal production. Given an arbitrary normal algorithm $\mathfrak{A}$, add on at the end of the schema for $\mathfrak{A}$ the new production $\Lambda \rightarrow \cdot\, \Lambda$, and denote by $\mathfrak{A}\cdot$ the normal algorithm determined by this enlarged schema. $\mathfrak{A}\cdot$ is closed, and $\mathfrak{A}\cdot$ is fully equivalent to $\mathfrak{A}$ relative to the alphabet of $\mathfrak{A}$.

Let us now show that the composition of two normal algorithms is again a normal algorithm. Let $\mathfrak{A}$ and $\mathfrak{B}$ be normal algorithms in an alphabet A. For each symbol b in A, form a new symbol $\overline{b}$, called the *correlate* of b. Let $\overline{A}$ be the alphabet consisting of the correlates of the symbols of A. We assume that A and $\overline{A}$ have no symbols in common. Let α and β be two symbols not in $A \cup \overline{A}$. Let $\mathfrak{S}_{\mathfrak{A}}$ be the schema of $\mathfrak{A}\cdot$ except that the terminal dot in terminal productions is replaced by α. Let $\mathfrak{S}_{\mathfrak{B}}$ be the schema of $\mathfrak{B}\cdot$ except that every symbol is replaced by its correlate, every terminal dot is replaced by β, productions of the form $\Lambda \rightarrow Q$ are replaced by $\alpha \rightarrow \alpha Q$, and productions $\Lambda \rightarrow \cdot\, Q$ are replaced by $\alpha \rightarrow \alpha\beta Q$. Consider the abbreviated schema

$$
\begin{array}{ll}
a\alpha \rightarrow \alpha a & (a \text{ in } A) \\
\alpha a \rightarrow \alpha \overline{a} & (a \text{ in } A) \\
\overline{a}\, b \rightarrow \overline{a}\, \overline{b} & (a, b \text{ in } A) \\
\overline{a}\, \beta \rightarrow \beta\, \overline{a} & (a \text{ in } A) \\
\beta\, \overline{a} \rightarrow \beta\, a & (a \text{ in } A) \\
a\, \overline{b} \rightarrow a\, b & (a, b \text{ in } A) \\
\alpha\beta \rightarrow \cdot\, \Lambda & \\
\quad \mathfrak{S}_{\mathfrak{B}} & \\
\quad \mathfrak{S}_{\mathfrak{A}} &
\end{array}
$$

This schema determines a normal algorithm $\mathfrak{G}$ over A such that $\mathfrak{G}(P) \approx \mathfrak{B}(\mathfrak{A}(P))$ for any word P in A. $\mathfrak{G}$ is called the *composition* of $\mathfrak{A}$ and $\mathfrak{B}$ and is denoted $\mathfrak{B} \circ \mathfrak{A}$. In general, by $\mathfrak{A}_n \circ \ldots \circ \mathfrak{A}_1$ we mean $\mathfrak{A}_n \circ (\ldots \circ (\mathfrak{A}_3 \circ (\mathfrak{A}_2 \circ \mathfrak{A}_1)) \ldots)$.

Let $\mathfrak{Y}$ be an algorithm in an alphabet A and let B be an extension of A. If we take a schema for $\mathfrak{Y}$ and prefix to it the production $b \rightarrow b$ for each symbol b in $B - A$, then the new schema determines a normal algorithm $\mathfrak{Y}_B$ in B such that $\mathfrak{Y}_B(P) \approx \mathfrak{Y}(P)$ for every word P in A, and $\mathfrak{Y}_B$ is not appli-

†In this and in all other definitions in this chapter, the existential quantifier 'there is' is meant in the ordinary 'classical' sense. When we assert that there exists an object of a certain kind, we do not necessarily imply that any human being has found or ever will find such an object. Thus, a function φ may be Markov-computable without our ever knowing it to be so.

cable to any word in B that contains any symbol of B − A. $\mathfrak{P}_B$ is fully equivalent to $\mathfrak{P}$ relative to A and is called the *propagation* of $\mathfrak{P}$ onto B.

Assume that $\mathfrak{A}$ is a normal algorithm in an alphabet A_1 and $\mathfrak{B}$ is a normal algorithm in an alphabet A_2. Let $A = A_1 \cup A_2$. Let $\mathfrak{A}_A$ and $\mathfrak{B}_A$ be the propagations of $\mathfrak{A}$ and $\mathfrak{B}$, respectively, onto A. Then the composition $\mathfrak{G}$ of $\mathfrak{A}_A$ and $\mathfrak{B}_A$ is called the *normal composition* of $\mathfrak{A}$ and $\mathfrak{B}$ and is denoted by $\mathfrak{B} \circ \mathfrak{A}$. (When $A_1 = A_2$, the normal composition of $\mathfrak{A}$ and $\mathfrak{B}$ is identical with the composition of $\mathfrak{A}$ and $\mathfrak{B}$; hence the notation $\mathfrak{B} \circ \mathfrak{A}$ is unambiguous.) $\mathfrak{G}$ is a normal algorithm over A such that $\mathfrak{G}(P) \approx \mathfrak{B}(\mathfrak{A}(P))$ for any word P in A_1, and $\mathfrak{G}$ is applicable to only those words P of A such that P is a word of A_1, $\mathfrak{A}$ is applicable to P, and $\mathfrak{B}$ is applicable to $\mathfrak{A}(P)$.

PROPOSITION 5.24

Let $\mathscr{T}$ be a Turing machine with alphabet A. Then there is a normal algorithm $\mathfrak{A}$ over A that is fully equivalent to the Turing algorithm $\mathrm{Alg}_{\mathscr{T}}$ relative to A.

Proof

Let $D = \{q_{k_0}, \ldots, q_{k_m}\}$, where $q_{k_0}, \ldots, q_{k_m}$ are the internal states of $\mathscr{T}$ and $q_{k_0} = q_0$. Write the algorithm schema for $\mathfrak{A}$ as follows: First, for all quadruples $q_j a_i a_k q_r$ of $\mathscr{T}$, take the production $q_j a_i \to q_r a_k$. Second, for each quadruple $q_j a_i L q_r$ of $\mathscr{T}$, take the productions $a_n q_j a_i \to q_r a_n a_i$ for all symbols a_n of A; then take the production $q_j a_i \to q_r a_0 a_i$. Third, for each quadruple $q_j a_i R q_r$ of $\mathscr{T}$, take the productions $q_j a_i a_n \to a_i q_r a_n$ for all symbols a_n of A; then take the production $q_j a_i \to a_i q_r a_0$. Fourth, write the productions $q_{k_i} \to \cdot \Lambda$ for each internal state q_{k_i} of $\mathscr{T}$, and finally take $\Lambda \to q_0$. This schema defines a normal algorithm $\mathfrak{A}$ over A, and it is easy to see that, for any word P of A, $\mathrm{Alg}_{\mathscr{T}}(P) \approx \mathfrak{A}(P)$.

COROLLARY 5.25

Every Turing-computable function is Markov-computable.

Proof

Let $f(x_1, \ldots, x_n)$ be standard Turing-computable by a Turing machine $\mathscr{T}$ with alphabet $A \supseteq \{|, B\}$. (Remember that B is a_0 and | is a_1.) We know that, for any natural numbers $k_1, \ldots, k_n$, if $f(k_1, \ldots, k_n)$ is not defined, then $\mathrm{Alg}_{\mathscr{T}}$ is not applicable to $\overline{(k_1, \ldots, k_n)}$, whereas, if $f(k_1, \ldots, k_n)$ is defined, then

$$\mathrm{Alg}_{\mathscr{F}}(\overline{(k_1,\ldots,k_n)}) \approx \mathrm{R}_1\overline{(k_1,\ldots,k_n)}\ \mathrm{B}\ \overline{f(k_1,\ldots,k_n)}\mathrm{R}_2$$

where R_1 and R_2 are (possibly empty) sequences of Bs. Let $\mathfrak{B}$ be a normal algorithm over A that is fully equivalent to $\mathrm{Alg}_{\mathscr{F}}$ relative to A. Let $\mathfrak{G}$ be the normal algorithm over $\{|, \mathrm{B}\}$ determined by the schema

$$\begin{aligned}
\alpha\mathrm{B} &\to \alpha\\
\alpha\,| &\to \beta\,|\\
\beta\,| &\to |\,\beta\\
\beta\mathrm{B} &\to \mathrm{B}\gamma\\
\gamma\,| &\to \beta\,|\\
\gamma\mathrm{B} &\to \gamma\\
\mathrm{B}\gamma &\to \cdot\,\Lambda\\
\beta &\to \cdot\,\Lambda\\
\Lambda &\to \alpha
\end{aligned}$$

If R_1 and R_2 are possibly empty sequences of Bs, then $\mathfrak{G}$, when applied to $\mathrm{R}_1\ \overline{(k_1,\ldots,k_n)}\ \mathrm{B}\ \overline{f(k_1,\ldots,k_n)}\ \mathrm{R}_2$, will erase R_1 and R_2. Finally, let $\mathfrak{A}_{n+1}^{n+1}$ be the normal 'projection' algorithm defined in Exercise 5.64(c). Then the normal composition $\mathfrak{A}_{n+1}^{n+1} \circ \mathfrak{G} \circ \mathfrak{B}$ is a normal algorithm that computes f.

Let $\mathfrak{A}$ be any algorithm over an alphabet $\mathrm{A} = \{\mathrm{a}_{j_0}, \ldots, \mathrm{a}_{j_m}\}$. We can associate with $\mathfrak{A}$ a partial number-theoretic function $\psi_{\mathfrak{A}}$ such that $\psi_{\mathfrak{A}}(n) = m$ if and only if either n is not the Gödel number† of a word of A and $m = 0$, or n and m are Gödel numbers of words P and Q of A such that $\mathfrak{A}(\mathrm{P}) = Q$.

PROPOSITION 5.26

If $\mathfrak{A}$ is a normal algorithm over $\mathrm{A} = \{\mathrm{a}_{j_0}, \ldots, \mathrm{a}_{j_m}\}$, then $\psi_{\mathfrak{A}}$ is partial recursive.

Proof

We may assume that the symbols of the alphabet of $\mathfrak{A}$ are of the form a_i. Given a simple production $\mathrm{P} \to \mathrm{Q}$, we call $2^1 3^{g(\mathrm{P})} 5^{g(\mathrm{Q})}$ its index; given a terminal production $\mathrm{P} \to \cdot\,\mathrm{Q}$, we let $2^2 3^{g(\mathrm{P})} 5^{g(\mathrm{Q})}$ be its index. If $\mathrm{P}_0 \to (\cdot)\mathrm{Q}_0, \ldots, \mathrm{P}_r \to (\cdot)\mathrm{Q}_r$ is an algorithm schema, we let its index be $2^{k_0} 3^{k_1} \ldots \mathrm{p}_r^{k_r}$, where k_i is the index of $\mathrm{P}_i \to (\cdot)\mathrm{Q}_i$. Let $\mathrm{Word}(u)$ be the recursive predicate that holds if and only if u is the Gödel number of a finite sequence of symbols of the form a_i:

†Here and below, we use the Gödel numbering of the language of Turing computability given in Section 5.3 (p. 321). Thus, the Gödel number $g(\mathrm{a}_i)$ of a_i is $7 + 4i$. In particular, $g(\mathrm{B}) = g(\mathrm{a}_0) = 7$ and $g(|) = g(\mathrm{a}_1) = 11$.

$$u \neq 0 \wedge [u = 1 \vee (\forall z)(z < \ell h(u) \Rightarrow (\exists y)(y < u \wedge (u)_z = 7 + 4y))]$$

Let SI(u) be the recursive predicate that holds when u is the index of a simple production: $\ell h(u) = 3 \wedge (u)_0 = 1 \wedge \text{Word}((u)_1) \wedge \text{Word}((u)_2)$. Similarly, TI($u$) is the recursive predicate that holds when u is the index of a terminal production: $\ell h(u) = 3 \wedge (u)_0 = 2 \wedge \text{Word}((u)_1) \wedge \text{Word}((u)_2)$. Let Ind($u$) be the recursive predicate that holds when u is the index of an algorithm schema: $u > 1 \wedge (\forall z)(z < \ell h(u) \Rightarrow \text{SI}((u)_z) \vee \text{TI}((u)_z))$. Let Lsub($x, y, e$) be the recursive predicate that holds if and only if e is the index of a production $\text{P} \rightarrow (\cdot)\text{Q}$ and x and y are Gödel numbers of words U and V such that P occurs in U, and V is the result of substituting Q for the leftmost occurrence of P in U:

$$\text{Word}(x) \wedge \text{Word}(y) \wedge (\text{SI}(e) \vee \text{TI}(e)) \wedge (\exists u)_{u \leqslant x}(\exists v)_{v \leqslant x}(x = u * (e)_1 * v \wedge y = u * (e)_2 * v \wedge \neg(\exists w)_{w \leqslant x}(\exists z)_{z \leqslant x}(x = w * (e)_1 * z \wedge w < u))$$

Let Occ(x, y) be the recursive predicate that holds when x and y are Gödel numbers of words U and V such that V occurs in U: $\text{Word}(x) \wedge \text{Word}(y) \wedge (\exists v)_{v \leqslant x}(\exists z)_{z \leqslant x}(x = v * y * z)$. Let End($e, z$) be the recursive predicate that holds when and only when z is the Gödel number of a word P, and e is the index of an algorithm schema defining an algorithm $\mathfrak{A}$ that cannot be applied to P (i.e., $\mathfrak{A} : \text{P}\sqsupset$):$\text{Ind}(e) \wedge \text{Word}(z) \wedge (\forall w)_{w < \ell h(e)} \neg\text{Occ}(z, ((e)_w)_1)$. Let SCons($e, y, x$) be the recursive predicate that holds if and only if e is the index of an algorithm schema and y and x are Gödel numbers of words V and U such that V arises from U by a simple production of the schema:

$$\text{Ind}(e) \wedge \text{Word}(x) \wedge \text{Word}(y) \wedge (\exists v)_{v < \ell h(e)}[\text{SI}((e)_v) \wedge \text{Lsub}(x, y, (e)_v) \wedge (\forall z)_{z < v} \neg\text{Occ}(x, ((e)_z)_1)]$$

Similarly, one defines the recursive predicate TCons(e, y, x), which differs from SCons(e, y, x) only in that the production in question is terminal. Let Der(e, x, y) be the recursive predicate that is true when and only when e is the index of an algorithm schema that determines an algorithm $\mathfrak{A}$, x is the Gödel number of a word U_0, y is the Gödel number of a sequence of words $\text{U}_0, \ldots, \text{U}_k (k \geq 0)$ such that, for $0 \leqslant i < k \dot{-} 1$, U_{i+1} arises from U_i by a production of the schema, and either $\mathfrak{A} : \text{U}_{k \dot{-} 1} \vdash \cdot\, \text{U}_k$ or $\mathfrak{A} : \text{U}_{k \dot{-} 1} \vdash \text{U}_k$ and $\mathfrak{A} : \text{U}_k \sqsupset$(or,if $k = 0$, just $\mathfrak{A} : \text{U}_k\sqsupset$):

$$\text{Ind}(e) \wedge \text{Word}(x) \wedge (\forall z)_{z < \ell h(y)} \text{Word}((y)_z) \wedge (y)_0 = x \wedge (\forall z)_{z < \ell h(y) \dot{-} 2} \text{SCons}(e, (y)_{z+1}, (y)_z) \wedge [(\ell h(y) = 1 \wedge \text{End}(e, (y)_0)) \vee (\ell h(y) > 1 \wedge \{\text{TCons}(e, (y)_{\ell h(y) \dot{-} 1}, (y)_{\ell h(y) \dot{-} 2}) \vee (\text{SCons}(e, (y)_{\ell h(y) \dot{-} 1}, (y)_{\ell h(y) \dot{-} 2}) \wedge \text{End}(e, (y)_{\ell h(y) \dot{-} 1}))\})]$$

Let $W_{\text{A}}(u)$ be the recursive predicate that holds if and only if u is the Gödel number of a word of A:

$$u \neq 0 \wedge (u = 1 \vee (\forall z)_{z < \ell h(u)}((u)_z = 7 + 4_{j_0} \vee \cdots \vee (u)_z = 7 + 4_{j_m})$$

Let e be the index of an algorithm schema for $\mathfrak{A}$. Now define the partial recursive function $\varphi(x) = \mu y((W_A(x) \wedge \mathrm{Der}(e,x,y)) \vee \neg W_A(x))$. But $\psi_{\mathfrak{A}}(x) = (\varphi(x))_{\ell\ell(\phi(x))\dot{-}1}$. Therefore, $\psi_{\mathfrak{A}}$ is partial recursive.

COROLLARY 5.27

Every Markov-computable function φ is partial recursive.

Proof

Let $\mathfrak{A}$ be a normal algorithm over $\{1, B\}$ such that $\varphi(k_1, \ldots, k_n) = l$ if and only if $\mathfrak{A}(\overline{(k_1, \ldots, k_n)}) = \bar{l}$. By Proposition 5.26, the function $\psi_{\mathfrak{A}}$ is partial recursive. Define the recursive function $\gamma(x) = \ell\ell(x) \dot{-} 1$. If $x = \Pi_{i=0}^{n} p_i^{11}$, then $n = \gamma(x)$. (Remember that a stroke |, which is an abbreviation for a_1, has Gödel number 11. So, if x is the Gödel number of the numeral $\bar{n}$, then $\gamma(x) = n$.) Let $\xi(k_1, \ldots, k_n)$ be the Gödel number of $\overline{(k_1, \ldots, k_n)}$:

$$\begin{aligned}\xi(k_1, \ldots, k_n) &= g(\overline{(k_1, \ldots, k_n)}) = g(|^{k_1+1}\mathrm{B}|^{k_2+1}\mathrm{B}\ldots\mathrm{B}|^{k_n+1}) \\ &= \Big(\prod_{i=0}^{k+1} p_i^{11}\Big) \cdot (p_{k_r+2})^7 \cdot \Big(\prod_{i=0}^{k_2+1} (p_{i+k_1+3})^{11}\Big) \cdot (p_{k_1+k_2+5})^7 \cdots \\ &\quad \cdot (p_{k_1+\cdots+k_n+2n\dot{-}3})^7 \cdot \Big(\prod_{i=0}^{k_n+1} (p_{i+k_1+\cdots+k_n+2n\dot{-}2})^{11}\Big)\end{aligned}$$

ξ is clearly recursive. Then $\varphi = \gamma \circ \psi_{\mathfrak{A}} \circ \xi$ is partial recursive.

The equivalence of Markov computability and Turing computability follows from Corollaries 5.25 and 5.27 and the known equivalence of Turing computability and partial recursiveness. Many other definitions of computability have been given, all of them turning out to be equivalent to Turing computability. One of the earliest definitions, λ-computability, was developed by Church and Kleene as part of the theory of λ-conversion (see Church, 1941). Its equivalence with the intuitive notion of computability is not immediately plausible and gained credence only when λ-computability was shown to be equivalent to partial recursiveness and Turing computability (see Kleene, 1936b; Turing, 1937). All reasonable variations of Turing computability seem to yield equivalent notions (see W. Oberschelp, 1958; Fischer, 1965).

5.6 DECISION PROBLEMS

A class of problems is said to be *unsolvable* if there is no effective procedure for solving each problem in the class. For example, given any polynomial $f(x)$ with integral coefficients (for example, $3x^5 - 4x^4 + 7x^2 - 13x + 12$), is

there an integer k such that $f(k) = 0$? We can certainly answer this question for various special polynomials, but is there a single general procedure that will solve the problem for every polynomial $f(x)$? (The answer is given below in paragraph 4.)

If we can arithmetize the formulation of a class of problems and assign to each problem a natural number, then this class is unsolvable if and only if there is no computable function h such that, if n is the number of a given problem, then $h(n)$ yields the solution of the problem. If Church's thesis is assumed, the function h has to be partial recursive, and we then have a more accessible mathematical question.

Davis (1977b) gives an excellent survey of research on unsolvable problems. Let us look at a few decision problems, some of which we already have solved.

1. Is a statement form of the propositional calculus a tautology? Truth tables provide an easy, effective procedure for answering any such question.

2. *Decidable and undecidable theories.* Is there a procedure for determining whether an arbitrary wf of a formal system $\mathscr{S}$ is a theorem of $\mathscr{S}$? If so, $\mathscr{S}$ is called *decidable*; otherwise, it is *undecidable.*

(a) The system L of Chapter 1 is decidable. The theorems of L are the tautologies, and we can apply the truth table method.
(b) The pure predicate calculus PP and the full predicate calculus PF were both shown to be recursively undecidable in Proposition 3.54.
(c) The theory RR and all its consistent extensions (including Peano arithmetic S) have been shown to be recursively undecidable in Corollary 3.46.
(d) The axiomatic set theory NBG and all its consistent extensions are recursively undecidable (see pages 269–70).
(e) Various theories concerning order structures or algebraic structures have been shown to be decidable (often by the method of quantifier elimination). Examples are the theory of unbounded densely ordered sets (see page 116 and Langford, 1927), the theory of abelian groups (Szmielew, 1955), and the theory of real closed fields (Tarski, 1951). For further information, consult Kreisel and Krivine (1967, Chap. 4); Chang and Keisler (1973, Chap. 1.5); Monk (1976, Chap. 13); Ershov *et al.* (1965); Rabin (1977); and Baudisch *et al.* (1985). On the other hand, the undecidability of many algebraic theories can be derived from the results in Chapter 3 (see Tarski, Mostowski and Robinson, 1953, II.6, III; Monk, 1976, Chap. 16).

3. *Logical validity.* Is a given wf of quantification theory logically valid? By Gödel's completeness theorem (Corollary 2.19), a wf is logically valid if and only if it is provable in the full predicate calculus PF. Since PF is recursively undecidable (Proposition 3.54), the problem of logical validity is recursively unsolvable.

However, there is a decision procedure for the logical validity of wfs of the pure monadic predicate calculus (Exercise 3.59).

There have been extensive investigations of decision procedures for various important subclasses of wfs of the pure predicate calculus; for example, the class $(\forall\ \exists\ \forall)$ of all closed wfs of the form $(\forall x)(\exists y)(\forall z)\mathscr{B}(x,y,z)$, where $\mathscr{B}(x,y,z)$ contains no quantifiers. See Ackermann (1954), Dreben and Goldfarb (1980) and Lewis (1979).

4. *Hilbert's Tenth Problem.* If $f(x_1, \ldots, x_n)$ is a polynomial with integral coefficients, are there integers $k_1, \ldots, k_n$ such that $f(k_1, \ldots, k_n) = 0$? This difficult decision problem is known as Hilbert's tenth problem.

For one variable, the solution is easy. When $a_0, a_1, \ldots, a_n$ are integers, any integer x such that $a_n^{x^n} + \ldots + a_1 x + a_0 = 0$ must be a divisor of a_0. Hence, when $a_0 \neq 0$, we can test each of the finite number of divisors of a_0. If $a_0 = 0$, then $x = 0$ is a solution. However, there is no analogous procedure when the polynomial has more than one variable. It was finally shown by Matiyasevich (1970) that there is no decision procedure for determining whether a polynomial with integral coefficients has a solution consisting of integers. His proof was based in part on some earlier work of Davis, Putnam and Robinson (1961). The proof ultimately relies on basic facts of recursion theory, particularly the existence of a non-recursive r.e. set (Proposition 5.21(e)). An up-to-date exposition may be found in Matiyasevich (1993).

5. *Word problems.*

(a) *Semi-Thue Systems.* Let $\mathrm{B} = \{\mathrm{b}_1, \ldots, \mathrm{b}_n\}$ be a finite alphabet. Remember that a word of B is a finite sequence of elements of B. Moreover, the empty sequence Λ is considered a word of B. By a *production* of B we mean an ordered pair $\langle \mathrm{u},\mathrm{v} \rangle$, where u and v are words of B. If $\mathrm{p} = \langle \mathrm{u},\mathrm{v} \rangle$ is a production of B, and if w and w′ are words of B, we write $\mathrm{w} \Rightarrow_{\mathrm{p}} \mathrm{w}'$ if w′ arises from w by replacing a part u of w by v. (Recall that u is a part of w if there exist (possibly empty) words w_1 and w_2 such that $\mathrm{w} = \mathrm{w}_1\mathrm{u}\mathrm{w}_2$.)

By a *semi-Thue system* on B we mean a finite set $\mathscr{S}$ of productions of B. For words w and w′ of B, we write $\mathrm{w} \Rightarrow_{\mathscr{S}} \mathrm{w}'$ if there is a finite sequence $\mathrm{w}_0, \mathrm{w}_1, \ldots, \mathrm{w}_k$ $(k \geq 0)$ of words of B such that $w = w_0, w' = w_k$, and, for $0 \leqslant i < k$, there is a production p of $\mathscr{S}$ such that $\mathrm{w}_i \Rightarrow_{\mathrm{p}} \mathrm{w}_{i+1}$. Observe that $\mathrm{w} \Rightarrow_{\mathscr{S}} \mathrm{w}$ for any word w of B. Moreover, if $\mathrm{w}_1 \Rightarrow_{\mathscr{S}} \mathrm{w}_2$ and $\mathrm{w}_2 \Rightarrow_{\mathscr{S}} \mathrm{w}_3$, then $\mathrm{w}_1 \Rightarrow_{\mathscr{S}} \mathrm{w}_3$. In addition, if $\mathrm{w}_1 \Rightarrow_{\mathscr{S}} \mathrm{w}_2$ and $\mathrm{w}_3 \Rightarrow_{\mathscr{S}} \mathrm{w}_4$, then $\mathrm{w}_1\mathrm{w}_3 \Rightarrow_{\mathscr{S}} \mathrm{w}_2\mathrm{w}_4$. Notice that there is no fixed order in which the productions have to be applied and that many different productions of $\mathscr{S}$ might be applicable to the same word.

By a *Thue system* we mean a semi-Thue system such that, for every production $\langle \mathrm{u},\mathrm{v} \rangle$, the *inverse* $\langle \mathrm{v},\mathrm{u} \rangle$ is also a production. Clearly, if $\mathscr{S}$ is a Thue system and $\mathrm{w} \Rightarrow_{\mathscr{S}} \mathrm{w}'$, then $\mathrm{w}' \Rightarrow_{\mathscr{S}} \mathrm{w}$. Hence, $\Rightarrow_{\mathscr{S}}$ is an equivalence relation on the set of words of the alphabet of $\mathscr{S}$.

Example
Let $\mathscr{S}^{\#}$ be the Thue system that has alphabet {b} and productions $\langle b^3, \Lambda\rangle$ and $\langle \Lambda, b^3\rangle$. It is easy to see that every word is transformable into b^2, b, or Λ.

By a *semigroup* we mean a non-empty set G together with a binary operation on G (denoted by the juxtaposition uv of elements u and v) that satisfies the associative law $x(yz) = (xy)z$. An element y such that $xy = yx = x$ for all x in G is called an *identity element*. If an identity element exists, it is unique and is denoted 1.

A Thue system $\mathscr{S}$ on an alphabet B determines a semigroup G with an identity element. In fact, for each word w of B, let [w] be the set of all words w′ such that $w \Rightarrow_{\mathscr{S}} w'$. [w] is just the equivalence class of w with respect to $\Rightarrow_{\mathscr{S}}$. Let G consist of the sets [w] for all words w of B. If U and V are elements of G, choose a word u in U and a word v in V. Let UV stand for the set [uv]. This defines an operation on G, since, if u′ is any word in U and v′ is any word in V, [uv] = [u′v′].

Exercises

5.65 For the set G determined by the Thue system $\mathscr{S}$, prove:
(a) (UV)W = U(VW) for all members U, V and W of G.
(b) The equivalence class [Λ] of the empty word Λ acts as an identity element of G.

5.66
(a) Show that a semigroup contains at most one identity element.
(b) Give an example of a semigroup without an identity element.

A Thue system $\mathscr{S}$ provides what is called a *finite presentation* of the corresponding semigroup G. The elements $b_1, \ldots, b_m$ of the alphabet of $\mathscr{S}$ are called *generators*, and the productions ⟨u,v⟩ of $\mathscr{S}$ are written in the form of equations u = v. These equations are called the *relations* of the presentation. Thus, in the example above, b is the only generator and $b^3 = \Lambda$ can be taken as the only relation. The corresponding semigroup is a cyclic group of order 3.

If $\mathscr{S}$ is a semi-Thue or Thue system, the *word problem* for $\mathscr{S}$ is the problem of determining, for any words w and w′, whether $w \Rightarrow_{\mathscr{S}} w'$.

Exercises

5.67 Show that, for the Thue system $\mathscr{S}^{\#}$ in the example, the word problem is solvable.

5.68 Consider the following Thue system $\mathscr{S}$. The alphabet is {a,b,c,d} and the productions are $\langle ac, \Lambda\rangle, \langle ca, \Lambda\rangle, \langle bd, \Lambda\rangle, \langle db, \Lambda\rangle, \langle a^3, \Lambda\rangle, \langle b^2, \Lambda\rangle, \langle ab, ba\rangle$, and their inverses.

(a) Show that $c \Rightarrow_{\mathscr{S}} a^2$ and $d \Rightarrow_{\mathscr{S}} b$.
(b) Show that every word of $\mathscr{S}$ can be transformed into one of the words a, a^2, b, ab, a^2b, and Λ.
(c) Show that the word problem for $\mathscr{S}$ is solvable. [*Hint*: To show that the six words of part (b) cannot be transformed into one another, use the cyclic group of order 6 generated by an element g, with $a = g^2$ and $b = g^3$.]

PROPOSITION 5.30

(Post, 1947) There exists a Thue system with a recursively unsolvable word problem.

Proof

Let $\mathscr{T}$ be a Turing machine with alphabet $\{a_0, a_1, \ldots, a_n\}$ and internal states $\{q_0, q_1, \ldots, q_m\}$. Remember that a tape description is a sequence of symbols describing the condition of $\mathscr{T}$ at any given moment; it consists of symbols of the alphabet of $\mathscr{T}$ plus one internal state q_j, and q_j is not the last symbol of the description. $\mathscr{T}$ is in state q_j, scanning the symbol following q_j, and the alphabet symbols, read from left to right, constitute the entire tape at the given moment. We shall construct a semi-Thue system $\mathscr{S}$ that will reflect the operation of $\mathscr{T}$: each action induced by quadruples of $\mathscr{T}$ will be copied by productions of $\mathscr{S}$. The alphabet of $\mathscr{S}$ consists of $\{a_0, a_1, \ldots, a_n, q_0, q_1, \ldots, q_m, \beta, \delta, \xi\}$. The symbol β will be placed at the beginning and end of a tape description in order to 'alert' the semi-Thue system when it is necessary to add an extra blank square on the left or right end of the tape. We wish to ensure that, if $W \underset{\mathscr{T}}{\mapsto} W'$, then $\beta W \beta \Rightarrow_{\mathscr{S}} \beta W' \beta$. The productions of $\mathscr{S}$ are constructed from the quadruples of $\mathscr{T}$ in the following manner.

(a) If $q_j a_i a_k q_r$ is a quadruple of $\mathscr{T}$, let $\langle q_j a_i, a_r q_k \rangle$ be a production of $\mathscr{S}$.
(b) If $q_j a_i R q_r$ is a quadruple of $\mathscr{T}$, let $\langle q_j a_i a_\ell, a_i q_r a_\ell \rangle$ be a production of $\mathscr{S}$ for every a_ℓ. In addition, let $\langle q_j a_i \beta a_i, q_r a_0 \beta \rangle$ be a production of $\mathscr{S}$. (This last production adds a blank square when $\mathscr{T}$ reaches the right end of the tape and is ordered to move right.)
(c) If $q_j a_i L q_r$ is a quadruple of $\mathscr{T}$, let $\langle a_\ell q_j a_i, q_r a_\ell a_i \rangle$ be a production of $\mathscr{S}$ for each a_ℓ. In addition, let $\langle \beta q_j a_i, \beta q_r a_0 a_i \rangle$ be a production of $\mathscr{S}$. (This last production adds a blank square to the left of the tape when this is required.)
(d) If there is no quadruple of $\mathscr{T}$ beginning with $q_j a_i$, let $\mathscr{S}$ contain the following productions: $\langle q_j a_i, \delta \rangle$, $\langle \delta a_\ell, \delta \rangle$ for all a_ℓ; $\langle \delta\ \beta, \xi \rangle$, $\langle a_\ell \xi, \xi \rangle$ for all a_ℓ; and $\langle \beta \xi, \xi \rangle$.

$\mathscr{T}$ stops when it is in a state q_j, scanning a symbol a_i, such that $q_j a_i$ does not begin a quadruple of $\mathscr{T}$. In such a case, $\mathscr{S}$ would replace $q_j a_i$ in the final tape description of $\mathscr{T}$ by δ. Then δ proceeds to annihilate all the other symbols to its right, including the rightmost β, whereupon it changes to ξ. ξ then annihilates all symbols to its left, including the remaining β. The final result is ξ alone. Hence:

($\square$) For any initial tape description α, $\mathscr{T}$ halts when and only when $\beta\alpha\beta \Rightarrow_{\mathscr{S}} \xi$

Now, enlarge $\mathscr{S}$ to a Thue system $\mathscr{S}'$ by adding to $\mathscr{S}$ the inverses of all the productions of $\mathscr{S}$. Let us show that:

(∇) For any initial tape description α of $\mathscr{T}$, $\beta\alpha\beta \Rightarrow_{\mathscr{S}'} \xi$ if and only if $\beta\alpha\beta \Rightarrow_{\mathscr{S}} \xi$

Clearly, if $\beta\alpha\beta \Rightarrow_{\mathscr{S}} \xi$, then $\beta\alpha\beta \Rightarrow_{\mathscr{S}'} \xi$. Conversely, assume for the sake of contradiction that $\beta\alpha\beta \Rightarrow_{\mathscr{S}'} \xi$, but it is not the case that $\beta\alpha\beta \Rightarrow_{\mathscr{S}} \xi$. Consider a sequence of words leading from $\beta\alpha\beta$ to ξ in $\mathscr{S}'$:

$$\beta\alpha\beta = w_0 \Rightarrow_{\mathscr{S}'} w_1 \Rightarrow_{\mathscr{S}'} \cdots \Rightarrow_{\mathscr{S}'} w_{t-1} \Rightarrow_{\mathscr{S}'} w_t = \xi$$

Here, each arrow is intended to indicate a single application of a production. It is clear from the definition of $\mathscr{S}$ that no production of $\mathscr{S}$ applies to ξ alone. Hence, the last step in the sequence $w_{t-1} \Rightarrow_{\mathscr{S}'} \xi$ must be the result of a production of $\mathscr{S}$. So, $w_{t-1} \Rightarrow_{\mathscr{S}} \xi$. Working backward, let us find the least p such that $w_p \Rightarrow_{\mathscr{S}} \xi$. Since we have assumed that it is not true that $\beta\alpha\beta \Rightarrow_{\mathscr{S}} \xi$, we must have $p > 0$. By the minimality of p, it is not true that $w_{p-1} \Rightarrow_{\mathscr{S}} w_p$. Therefore, $w_p \Rightarrow_{\mathscr{S}} w_{p-1}$. Examination of the productions of $\mathscr{S}$ shows that each of the words $w_0, w_1, \ldots, w_t$ must contain exactly one of the symbols $q_0, q_1, \ldots, q_m, \delta$, or ξ, and that, to such a word, at most one production of $\mathscr{S}$ is applicable. But, w_p is transformed into both w_{p+1} and w_{p-1} by productions of $\mathscr{S}$. Hence, $w_{p-1} = w_{p+1}$. But, $w_{p+1} \Rightarrow_{\mathscr{S}} \xi$. Hence, $w_{p-1} \Rightarrow_{\mathscr{S}} \xi$, contradicting the definition of p. This establishes (∇).

Now, let $\mathscr{T}$ be a Turing machine with a recursively unsolvable halting problem (Proposition 5.14). Construct the corresponding Thue system $\mathscr{S}'$ as above. Then, by ($\square$) and (∇), for any tape description α, $\mathscr{T}$ halts if and only if $\beta\alpha\beta \Rightarrow_{\mathscr{S}'} \xi$. So, if the word problem for $\mathscr{S}'$ were recursively solvable, the halting problem for $\mathscr{T}$ would be recursively solvable. (The function that assigns to the Gödel number of α the Gödel number of $\langle\beta\alpha\beta, \xi\rangle$ is clearly recursive under a suitable arithmetization of the symbolism of Turing machines and Thue systems.) Thus, $\mathscr{S}'$ has a recursively unsolvable word problem.

That the word problem is unsolvable even for certain Thue systems on a *two-element* alphabet (semigroups with two generators) was proved by Hall (1949).

(b) *Finitely presented groups.* A *finite presentation* of a group consists of a finite set of generators $g_1, \ldots, g_r$ and a finite set of equations $W_1 = W_1', \ldots, W_t = W_t'$ between words of the alphabet $B = \{g_1, \ldots,$

$g_r, g_1^{-1}, \ldots, g_r^{-1}\}$. What is really involved here is a Thue system $\mathscr{S}$ with alphabet B, productions $\langle W_1, W_1' \rangle, \ldots, \langle W_t, W_t' \rangle$ and their inverses, and all the productions $\langle g_i g_i^{-1}, \Lambda \rangle, \langle g_i^{-1} g_i, \Lambda \rangle$ and their inverses. The corresponding semigroup G is actually a group and is called a *finitely presented group*. The word problem for G (or, rather, for the finite presentation of G) is the word problem for the Thue system $\mathscr{S}$.

Problems that concern word problems for finitely presented groups are generally much more difficult than corresponding problems for finitely presented semigroups (Thue systems). The existence of a finitely presented group with a recursively unsolvable word problem was proved, independently, by Novikov (1955) and Boone (1959). Other proofs have been given by Higman (1961), Britton (1963), and McKenzie and Thompson (1973). (See also Rotman, 1973.) Results on other decision problems connected with groups may be found in Rabin (1958). For corresponding problems in general algebraic systems, consult Evans (1951).

Appendix
Second-Order Logic

Our treatment of quantification theory in Chapter 2 was confined to first-order logic, that is, the variables used in quantifiers were only individual variables. The axiom systems for formal number theory in Chapter 3 and set theory in Chapter 4 also were formulated within first-order languages. This restriction brings with it certain advantages and disadvantages, and we wish now to see what happens when the restriction is lifted. That will mean allowing quantification with respect to predicate and function variables. Emphasis will be on second-order logic, since the important differences between first-order and higher-order logics already reveal themselves at the second-order level. Our treatment will offer only a sketch of the basic ideas and results of second-order logic.

Let L1C be the first-order language in which C is the set of non-logical constants (that is, individual constants, function letters, and predicate letters). Start with the language L1C, and add function variables $\mathbf{g}_i^n$ and predicate variables $\mathbf{R}_i^n$, where n and i are any positive integers.[†] (We shall use $\mathbf{g}^n, \mathbf{h}^n, \ldots$ to stand for any function variables of n arguments, and $\mathbf{R}^n, \mathbf{S}^n, \ldots, \mathbf{X}^n, \mathbf{Y}^n, \mathbf{Z}^n$ to stand for any predicate variables of n arguments. We shall also omit the superscript n when the value of n is clear from the context.) Let $\langle u \rangle_n$ stand for any sequence of individual variables $u_1, \ldots, u_n$[‡] and let $\forall \langle u \rangle_n$ stand for the expression $(\forall u_1) \ldots (\forall u_n)$. Similarly, let $\langle t \rangle_n$ stand for a sequence of terms $t_1, \ldots, t_n$. We expand the set of terms by allowing formation of terms $\mathbf{g}^n(\langle t \rangle_n)$, where $\mathbf{g}^n$ is a function variable, and we then expand the set of formulas by allowing formation of atomic formulas

[†]We use bold letters to avoid confusion with function letters and predicate letters. Note that function letters and predicate letters are supposed to denote specific operations and relations, whereas function variables and predicate variables vary over arbitrary operations and relations.

[‡]In particular, $\langle x \rangle_n$ will stand for $x_1, \ldots, x_n$.

$A_i^n(\langle t\rangle_n)$ and $\mathbf{R}^n(\langle t\rangle_n)$ where $\langle t\rangle_n$ is any sequence of the newly enlarged set of terms, A_i^n is any predicate letter of C and $\mathbf{R}^n$ is any n-ary predicate variable. Finally, we expand the set of formulas by quantification $(\forall \mathbf{g}^n)\mathscr{B}$ and $(\forall \mathbf{R}^n)\mathscr{B}$ with respect to function and predicate variables.

Let L2C denote the *second-order* language obtained in this way. The language L2C will be called a *full* second-order language. The adjective 'full' indicates that we allow both function variables and predicate variables and that there is no restriction on the arity n of those variables. An example of a non-full second-order language is the second-order monadic predicate language in which there are no function letters or variables, no predicate letters, and only monadic predicate variables.†

It is not necessary to take $=$ as a primitive symbol, since it can be defined in the following manner.

DEFINITIONS

$$t = u \text{ stands for } (\forall \mathbf{R}^1)(\mathbf{R}^1 t \Leftrightarrow \mathbf{R}^1 u)$$
$$\mathbf{g}^n = \mathbf{h}^n \text{ stands for } \forall\langle x\rangle_n(\mathbf{g}^n(\langle x\rangle_n) = \mathbf{h}^n(\langle x\rangle_n))$$
$$\mathbf{R}^n = \mathbf{S}^n \text{ stands for } \forall\langle x\rangle_n(\mathbf{R}^n(\langle x\rangle_n) \Leftrightarrow \mathbf{S}^n(\langle x\rangle_n))$$

STANDARD SECOND-ORDER SEMANTICS FOR L2C

For a given language L2C, let us start with a first-order interpretation with domain D. In the first-order case, we defined satisfaction for the set $\sum$ of denumerable sequences of members of D. Now, instead of $\sum$, we use the set $\sum_2$ of functions s that assign to each individual variable a member of D, to each function variable $\mathbf{g}^n$ some n-ary operation $s(\mathbf{g}^n)$ on D, and to each predicate variable $\mathbf{R}^n$ some n-ary relation‡ $s(\mathbf{R}^n)$ on D. For each such s, we extend the denotations determined by s by specifying that, for any terms $t_1, \ldots, t_n$ and any function variable $\mathbf{g}^n$, the denotation $s(\mathbf{g}^n(t_1, \ldots, t_n))$ is $s(\mathbf{g}^n)(s(t_1), \ldots, s(t_n))$. The first-order definition of satisfaction is extended as follows:

(a) For any predicate variable $\mathbf{R}^n$ and any finite sequence $\langle t\rangle_n$ of terms, s satisfies $\mathbf{R}^n(\langle t\rangle_n)$ if and only if $\langle s(t_1), \ldots, s(t_n)\rangle \in s(\mathbf{R}^n)$;

†Third-order logics are obtained by adding function and predicate letters and variables that can have as arguments individual variables, function and predicate letters, and second-order function and predicate variables, and then allowing quantification with respect to the new function and predicate variables. This procedure can be iterated to obtain nth-order logics for all $n \geq 1$.

‡An n-ary relation on D is a subset of the set D^n of n-tuples of D. When $n = 1$, an n-ary relation is just a subset of D.

(b) s satisfies $(\forall \mathbf{g}^n)\mathscr{B}$ if and only if s' satisfies $\mathscr{B}$ for every s' in $\sum_2$ that agrees with s except possibly at $\mathbf{g}^n$;

(c) s satisfies $(\forall \mathbf{R}^n)\mathscr{B}$ if and only if s' satisfies $\mathscr{B}$ for every s' in $\sum_2$ that agrees with s except possibly at $\mathbf{R}^n$.

The resulting interpretation $\mathscr{A}$ is called a *standard interpretation* of the given language.

A formula $\mathscr{B}$ is said to be *true* for a standard interpretation $\mathscr{A}$ (written $\mathscr{A} \models \mathscr{B}$) if $\mathscr{B}$ is satisfied by every s in $\sum_2$. $\mathscr{B}$ is *false* for $\mathscr{A}$ if no sequence s in $\sum_2$ satisfies $\mathscr{B}$.

A formula $\mathscr{B}$ is said to be *standardly valid* if $\mathscr{B}$ is true for all standard interpretations. $\mathscr{B}$ is said to be *standardly satisfiable* if $\mathscr{B}$ is satisfied by some s in $\sum_2$ in some standard interpretation. A formula $\mathscr{C}$ is said to be a *standard logical consequence* of a set Γ of formulas if, for every standard interpretation, every s in $\sum_2$ that satisfies every formula in Γ also satisfies $\mathscr{C}$. A formula $\mathscr{B}$ is said to *standardly logically imply* a formula $\mathscr{C}$ if $\mathscr{C}$ is a logical consequence of $\{\mathscr{B}\}$.

The basic properties of satisfaction, truth, logical consequence, and logical implication that held in the first-order case (see (I)–(XI) on pp. 61–3) also hold here for their standard versions. In particular, a sentence $\mathscr{B}$ is standardly satisfiable if and only if $\mathscr{B}$ is true for some standard interpretation.

We shall see that second-order languages have much greater expressive power than first-order languages. This is true even in the case where the set C of non-logical constants is empty. The corresponding language L2$\emptyset$ will be denoted L2 and called the *pure* full second-order language. Consider the following sentence in L2.

$$(1) \qquad (\exists \mathbf{g})(\exists x)(\forall \mathbf{R})[(\mathbf{R}(x) \wedge (\forall y)(\mathbf{R}(y) \Rightarrow \mathbf{R}(\mathbf{g}(y)))) \Rightarrow (\forall x)\mathbf{R}(x)]$$

This sentence is true for a standard interpretation if and only if the domain D is finite or denumerable. To see this, consider an operation g and element x given by this sentence. By induction, define the sequence x, $g(x)$, $g(g(x))$, $g(g(g(x)))$, ..., and let R be the set of objects in this sequence. R is finite or denumerable, and (1) tells us that every object in D is in R. Hence, $D = R$ and D is finite or denumerable. Conversely, assume that D is finite or denumerable. Let F be a one–one function from D onto ω (when D is denumerable) or onto an initial segment $\{0, 1, \ldots, n\}$ of ω (when D is finite).† Let $x = F^1(0)$ and define an operation g on D in the following manner. When D is denumerable, $g(u) = F^{-1}(F(u) + 1)$ for all u in D; when D is finite, let $g(u) = F^{-1}(F(u) + 1))$ if $F(u) < n$ and $g(u) = x$ if $F(u) = n$. With this choice of g and x, (1) holds.

†Remember that the domain of an interpretation is assumed to be non-empty.

Exercise

A.1 Show that there is no first-order sentence $\mathscr{B}$ such that $\mathscr{B}$ is true in an interpretation if and only if its domain is finite or denumerable. [Hint: Use Corollary 2.22.]

Let us introduce the abbreviations $\mathbf{Y}^1 \subseteq \mathbf{X}^1$ for $(\forall u)(\mathbf{Y}^1(u) \Rightarrow \mathbf{X}^1(u))$, $\text{NonEm}(\mathbf{X}^1)$ for $(\exists u)(\mathbf{X}^1(u))$, and $\text{Asym}(\mathbf{R}^2, \mathbf{X}^1)$ for $(\forall u)(\forall v)(\mathbf{X}^1(u) \wedge \mathbf{X}^1(v) \wedge \mathbf{R}^2(u, v) \Rightarrow \neg\mathbf{R}^2(v, u))$. Let $\mathbf{R}^2$ We $\mathbf{X}^1$ stand for the second-order formula

$$\text{Asym}(\mathbf{R}^2, \mathbf{X}^1) \wedge (\forall \mathbf{Y}^1)(\mathbf{Y}^1 \subseteq \mathbf{X}^1 \wedge \text{NonEm}(\mathbf{Y}^1) \Rightarrow (\exists u)(\mathbf{Y}^1(u) \wedge (\forall v)(\mathbf{Y}^1(v) \wedge v \neq u \Rightarrow \mathbf{R}^2(u, v))))$$

Then $\mathbf{R}^2$ We $\mathbf{X}^1$ is satisfied by an assignment in a given standard interpretation if and only if the binary relation assigned to $\mathbf{R}^2$ well-orders the set assigned to $\mathbf{X}^1$.

Let $\text{Suc}(u, v, \mathbf{R}^2)$ stand for $\mathbf{R}^2(v, u) \wedge (\forall w)\neg(\mathbf{R}^2(v, w) \wedge \mathbf{R}^2(w, u))$, and let $\text{First}(u, \mathbf{R}^2)$ stand for $(\forall v)(v \neq u \Rightarrow \mathbf{R}^2(u, v))$. Consider the following second-order formula.

$$(2) \qquad (\exists \mathbf{R}^2)(\exists \mathbf{X}^1)(\mathbf{R}^2 \text{ We } \mathbf{X}^1 \wedge (\forall u)\mathbf{X}^1(u) \wedge (\forall u)(\neg\text{First}(u, \mathbf{R}^2) \Rightarrow (\exists v)\text{Suc}(u, v, \mathbf{R}^2)) \wedge (\exists u)(\forall v)(v \neq u \Rightarrow \mathbf{R}^2(v, u)))$$

This is true for a standard interpretation if and only if there is a well-ordering of the domain in which every element other than the first is a successor and there is a last element. But this is equivalent to the domain being finite. Hence, (2) is true for a standard interpretation if and only if its domain is finite.

Exercise

A.2 (a) Show that, for every natural number n, there is a first-order sentence the models of which are all interpretations whose domain contains at least n elements. (b) Show that, for every positive integer n, there is a first-order theory the models of which are all interpretations whose domain contains exactly n elements. (c) Show that there is no first-order sentence $\mathscr{B}$ that is true for any interpretation if and only if its domain is finite.

The second-order sentence (1) $\wedge \neg$(2) is true for a standard interpretation if and only if the domain is denumerable.

Exercises

A.3 Show that there is no first-order sentence $\mathscr{B}$ the models of which are all interpretations whose domain is denumerable.

A.4 Construct a second-order formula $\mathrm{Den}(\mathbf{X}^1)$ that is satisfied by an assignment in a standard interpretation if and only if the set assigned to $\mathbf{X}^1$ is denumerable.

SECOND-ORDER THEORIES

We define a *second-order theory* in a language L2C by adding the following new logical axioms and rules to the first-order axioms and rules.

(B4a) $(\forall \mathbf{R}^n)\mathscr{B}(\mathbf{R}^n) \Rightarrow \mathscr{B}(\mathbf{W}^n)$, where $\mathscr{B}(\mathbf{W}^n)$ arises from $\mathscr{B}(\mathbf{R}^n)$ by replacing all free occurrences of $\mathbf{R}^n$ by $\mathbf{W}^n$ and $\mathbf{W}^n$ is free for $\mathbf{R}^n$ in $\mathscr{B}(\mathbf{R}^n)$.

(B4b) $(\forall \mathbf{g}^n)\mathscr{B}(\mathbf{g}^n) \Rightarrow \mathscr{B}(\mathbf{h}^n)$, where $\mathscr{B}(\mathbf{h}^n)$ arises from $\mathscr{B}(\mathbf{g}^n)$ by replacing all free occurrences of $\mathbf{g}^n$ by $\mathbf{h}^n$ and $\mathbf{h}^n$ is free for $\mathbf{g}^n$ in $\mathscr{B}(\mathbf{g}^n)$.

(B5a) $(\forall \mathbf{R}^n)(\mathscr{B} \Rightarrow \mathscr{C}) \Rightarrow (\mathscr{B} \Rightarrow (\forall \mathbf{R}^n)\mathscr{C})$, where $\mathbf{R}^n$ is not free in $\mathscr{B}$.

(B5b) $(\forall \mathbf{g}^n)(\mathscr{B} \Rightarrow \mathscr{C}) \Rightarrow (\mathscr{B} \Rightarrow (\forall \mathbf{g}^n)\mathscr{C})$, where $\mathbf{g}^n$ is not free in $\mathscr{B}$.

COMPREHENSION SCHEMA (COMP)

$(\exists \mathbf{R}^n)(\forall \langle x \rangle_n)(\mathbf{R}^n(\langle x \rangle_n) \Leftrightarrow \mathscr{B})$, provided that all free variables of $\mathscr{B}$ occur in $\langle x \rangle_n$ and $\mathbf{R}^n$ is not free in $\mathscr{B}$.

FUNCTION DEFINITION SCHEMA (FUNDEF)

$$(\forall \mathbf{R}^{n+1})[(\forall \langle x \rangle_n)(\exists_1 y)\mathbf{R}^{n+1}(\langle x \rangle_n, y) \Rightarrow (\exists \mathbf{g}^n)(\forall \langle x \rangle_n)\mathbf{R}^{n+1}(\langle x \rangle_n, \mathbf{g}^n(\langle x \rangle_n))]$$

NEW RULES

(Gen2a) $(\forall \mathbf{R}^n)\mathscr{B}$ follows from $\mathscr{B}$

(Gen2b) $(\forall \mathbf{g}^n)\mathscr{B}$ follows from $\mathscr{B}$

Exercises

A.5 Show that we can prove analogues of the usual equality axioms (A6)–(A7) in any second-order theory:

(i) $\vdash t = t \wedge \mathbf{g}^n = \mathbf{g}^n \wedge \mathbf{R}^n = \mathbf{R}^n$

(ii) $\vdash t = s \Rightarrow (\mathscr{B}(t, t) \Rightarrow \mathscr{B}(t, s))$, where $\mathscr{B}(t, s)$ arises from $\mathscr{B}(t, t)$ by replacing zero or more occurrences of t by s, provided that s is free for t in $\mathscr{B}(t, t)$.

(iii) $\vdash \mathbf{g}^n = \mathbf{h}^n \Rightarrow (\mathscr{B}(\mathbf{g}^n, \mathbf{g}^n) \Rightarrow \mathscr{B}(\mathbf{g}^n, \mathbf{h}^n))$, where $\mathscr{B}(\mathbf{g}^n, \mathbf{h}^n)$ arises from $\mathscr{B}(\mathbf{g}^n, \mathbf{g}^n)$ by replacing zero or more occurrences of $\mathbf{g}^n$ by $\mathbf{h}^n$, provided that $\mathbf{h}^n$ is free for $\mathbf{g}^n$ in $\mathscr{B}(\mathbf{g}^n, \mathbf{g}^n)$.

(iv) $\vdash \mathbf{R}^n = \mathbf{S}^n \Rightarrow (\mathscr{B}(\mathbf{R}^n, \mathbf{R}^n) \Rightarrow \mathscr{B}(\mathbf{R}^n, \mathbf{S}^n))$, where $\mathscr{B}(\mathbf{R}^n, \mathbf{S}^n)$ arises from $\mathscr{B}(\mathbf{R}^n, \mathbf{R}^n)$ by replacing zero or more occurrences of $\mathbf{R}^n$ by $\mathbf{S}^n$, provided that $\mathbf{S}^n$ is free for $\mathbf{R}^n$ in $\mathscr{B}(\mathbf{R}^n, \mathbf{R}^n)$.

A.6 Formulate and prove a second-order analogue of the first-order deduction theorem (Proposition 2.5).

Let PC2 denote the second-order theory in the language L2*C* without any non-logical axioms. PC2 is called a *second-order predicate calculus*.

PROPOSITION A.1 (SOUNDNESS)

Every theorem of PC2 is standardly valid.

Proof

That all the logical axioms (except Comp and FunDef) are standardly valid and that the rules of inference preserve standard validity follow by arguments like those for the analogous first-order properties. The standard validity of Comp and FunDef follows by simple set-theoretic arguments.

We shall see that the converse of Proposition A.1 does not hold. This will turn out to be not a consequence of a poor choice of axioms and rules, but an inherent incompleteness of second-order logic.

Let us consider the system of natural numbers. No first-order theory will have as its models those and only those interpretations that are isomorphic to the system of natural numbers.† However, a second-order characterization of the natural numbers is possible. Let AR2 be the conjunction of the axioms (S1)–(S8) of the theory S of formal arithmetic (see p. 155), and the following second-order principle of mathematical induction:

$$(2S9) \quad (\forall R^1)[R^1(0) \wedge (\forall x)(R^1(x) \Rightarrow R^1(x')) \Rightarrow (\forall x)R^1(x)]$$

Notice that, with the help of (Comp), all instances of the first-order axiom schema (S9) can be derived from (2S9).‡

†Let K be any first-order theory in the language of arithmetic whose axioms are true in the system of natural numbers. Add a new individual constant b and the axioms $b \neq \bar{n}$ for every natural number n. The new theory K* is consistent, since any finite set of its axioms has a model in the system of natural numbers. By Proposition 2.17, K* has a model, but that model cannot be isomorphic to the system of natural numbers, since the object denoted by b cannot correspond to a natural number.

‡In AR2, the function letters for addition and multiplication and the associated axioms (S5)–(S8) can be omitted. The existence of operations satisfying (S5)–(S8) can then be proved. See Mendelson (1973, Sections 2.3 and 2.5).

For any standard interpretation that is a model of AR2 we can prove the following result that justifies inductive definition.

PROPOSITION A.2 (ITERATION THEOREM)

Let $\mathcal{M}$ be a standard interpretation that is a model of AR2, and let D be the domain of $\mathcal{M}$. Let c be an element of an arbitrary set W and let g be a singulary operation of W. Then there is a unique function F from D into W such that $D(0) = c$ and $(\forall x)(x \in D \Rightarrow F(x') = g(F(x)))$.†

Proof

Let $\mathscr{C}$ be the set of all subsets H of $D \times W$ such that $\langle 1, c\rangle \in H$ and $(\forall x)(\forall w)(\langle x, w\rangle \in H \Rightarrow \langle x', g(w)\rangle \in H)$. Note that $D \times W \in \mathscr{C}$. Let F be the intersection of all sets H in $\mathscr{C}$. We leave it to the reader to prove the following assertions:

(a) $F \in \mathscr{C}$
(b) F is a function from D into W. [Hint: Let B be the set of all x in D for which there is a unique w in W such that $\langle x, w\rangle \in F$. By mathematical induction, show that $B = D$.]
(c) $F(1) = c$.
(d) $F(x') = g(F(x))$ for all x in D.

The uniqueness of F can be shown by a simple application of mathematical induction.

PROPOSITION A.3 (CATEGORICITY OF AR2)

Any two standard interpretations $\mathcal{M}$ and $\mathcal{M}^*$ that are models of AR2 are isomorphic.

Proof

Let D and D^* be the domains of $\mathcal{M}$ and $\mathcal{M}^*$, 0 and 0^* the respective zero elements, and f and f^* the respective successor operations. By the iteration theorem applied to $\mathcal{M}$, with $W = D^*$, $c = 0^*$ and $g = f^*$, we obtain a function F from D into D^* such that $F(0) = 0^*$ and $F(f(x)) = f^*(F(x))$ for any x in D. An easy application of mathematical induction in $\mathcal{M}^*$ shows that every element of D^* is in the range of F. To show that F is one–one, apply

†In order to avoid cumbersome notation, '0' denotes the interpretation in $\mathcal{M}$ of the individual constant '0', and 'x'' denotes the result of the application to the object x of the interpretation of the successor function.

mathematical induction in $\mathscr{M}$ to the set of all x in D such that $(\forall y)[(y \in D \wedge y \neq x) \Rightarrow F(x) \neq F(y)]$.†

Let $\mathscr{A}$ consist of the non-logical constants of formal arithmetic (zero, successor, addition, multiplication, equality). Let $\mathscr{N}$ be the standard interpretation of L2$\mathscr{A}$ with the set of natural numbers as its domain and the usual interpretations of the non-logical constants.

PROPOSITION A.4

Let $\mathscr{B}$ be any formula of L2$\mathscr{A}$. Then $\mathscr{B}$ is true in $\mathscr{N}$ if and only if AR2 $\Rightarrow \mathscr{B}$ is standardly valid.

Proof

Assume AR2 $\Rightarrow \mathscr{B}$ is standardly valid. So, AR2 $\Rightarrow \mathscr{B}$ is true in $\mathscr{N}$. But AR2 is true in $\mathscr{N}$. Hence, $\mathscr{B}$ is true in $\mathscr{N}$. Conversely, assume $\mathscr{B}$ is true in $\mathscr{N}$. We must show that AR2 $\Rightarrow \mathscr{B}$ is standardly valid. Assume that AR2 is true in some standard interpretation $\mathscr{M}$ of L2$\mathscr{A}$. By the categoricity of AR2, $\mathscr{M}$ is isomorphic to $\mathscr{N}$. Therefore, since $\mathscr{B}$ is true in $\mathscr{N}$, $\mathscr{B}$ is true in $\mathscr{M}$. Thus, AR2 $\Rightarrow \mathscr{B}$ is true in every standard interpretation of L2$\mathscr{A}$, that is, AR2 $\Rightarrow \mathscr{B}$ is standardly valid.

PROPOSITION A.5

(a) The set SV of standardly valid formulas of L2$\mathscr{A}$ is not effectively enumerable.
(b) SV is not recursively enumerable, that is, the set of Gödel numbers of formulas in SV is not recursively enumerable.

Proof

(a) Assume that SV is effectively enumerable. Then, by Proposition A4, we could effectively enumerate the set $\mathscr{TR}$ of all true formulas of first-order arithmetic by running through SV, finding all formulas of the form AR2 $\Rightarrow \mathscr{B}$, where $\mathscr{B}$ is a formula of first-order arithmetic, and listing those formulas $\mathscr{B}$. Then the theory $\mathscr{TR}$ would be decidable, since, for any closed formula $\mathscr{C}$, we could effectively enumerate $\mathscr{TR}$ until either $\mathscr{C}$ or its negation appears. By Church's thesis, $\mathscr{TR}$ would be recursively decidable, contradicting Corollary 3.46 (since $\mathscr{TR}$ is a consistent extension of RR).
(b) This follows from part (a) by Church's thesis.

†Details of the proof may be found in Mendelson (1973, Section 2.7).

The use of Church's thesis in the proof could be avoided by a consistent use of recursion-theoretic language and results. The same technique as the one used in part (a), together with Tarski's theorem (Corollary 3.44), would show the stronger result that the set (of Gödel numbers) of the formulas in SV is not arithmetical.

COROLLARY A.6

The set of all standardly valid formulas is not effectively (or recursively) enumerable.

Proof

An enumeration of all standardly valid formulas would yield an enumeration of all standardly valid formulas of L2$\mathscr{A}$, since the set of formulas of L2$\mathscr{A}$ is decidable (recursively decidable).

COROLLARY A.7

There is no axiomatic formal system whose theorems are the standardly valid formulas of L2$\mathscr{A}$.

Proof

If there were such an axiom system, we could enumerate the standardly valid formulas of L2$\mathscr{A}$, contradicting Corollary A.5.

PROPOSITION A.8 (INCOMPLETENESS OF STANDARD SEMANTICS)

There is no axiomatic formal system whose theorems are all standardly valid formulas.

Proof

If there were such an axiom system, we could enumerate the set of all standardly valid formulas, contradicting Corollary A.6.

Proposition A.8 sharply distinguishes second-order logic from first-order logic, since Gödel's completeness theorem tells us that there is an axiomatic formal system whose theorems are all logically valid first-order formulas.

Here are some additional important properties enjoyed by first-order theories that do not hold for second-order theories.

(I) Every consistent theory has a model. To see that this does not hold for second-order logic (with 'model' meaning 'model in the sense of the standard semantics'), add to the theory AR2 a new individual constant b. Let $\mathscr{T}$ be the theory obtained by adding to AR2 the set of axioms $b \neq \bar{n}$ for all natural number n. $\mathscr{T}$ is consistent. (Any proof involves a finite number of the axioms $b \neq \bar{n}$. AR2 plus any finite number of the axioms $b \neq \bar{n}$ has the standard interpretation as a model, with b interpreted as a suitable natural number. So, every step of the proof would be true in $\mathscr{N}$. Therefore, a contradiction cannot be proved.). But $\mathscr{T}$ has no standard model. (If $\mathscr{M}$ were such a model, AR2 would be true in $\mathscr{M}$. Hence, $\mathscr{M}$ would be isomorphic to $\mathscr{N}$ and so, the domain of $\mathscr{M}$ would consist of the objects denoted by the numerals $\bar{n}$. But this contradicts the requirement that the domain of $\mathscr{M}$ would have to have an object denoted by 'b' that would satisfy the axioms $b \neq \bar{n}$ for all natural numbers n.)

(II) The compactness property: a set Γ of formulas has a model if and only if every finite subset of Γ has a model. A counterexample is furnished by the set of axioms of the theory $\mathscr{T}$ in (I) above.

(III) The upward Skolem–Löwenheim theorem: every theory that has an infinite model has models of every infinite cardinality. In second-order logic this fails for the theory AR2. By Proposition A.3, all models of AR must be denumerable.

(IV) The downward Skolem–Löwenheim theorem: every model $\mathscr{M}$ of a theory has a countable elementary submodel†. In second-order logic, a counterexample is furnished by the second-order categorical theory for the real number system.‡ Another argument can be given by the following considerations. We can express by the following second-order formula $\mathscr{P}(\mathbf{Y}^1, \mathbf{X}^1)$ the assertion that $\mathbf{Y}^1$ is equinumerous with the power set of $\mathbf{X}^1$:

$$(\exists \mathbf{R}^2)[(\forall x_1)(\forall x_2)(\mathbf{X}^1(x_1) \wedge \mathbf{X}^1(x_2) \wedge (\forall y)(\mathbf{Y}^1(y) \Rightarrow [R^2(x_1, y) \Leftrightarrow \mathbf{R}^2(x_2, y)]) \Rightarrow x_1 = x_2) \wedge (\forall \mathbf{W}^1)(\mathbf{W}^1 \subseteq \mathbf{Y}^1 \Rightarrow (\exists x)(\mathbf{X}^1(x) \wedge (\forall y)(\mathbf{W}^1(y) \Leftrightarrow \mathbf{R}^2(x, y))))]$$

$\mathbf{R}^2$ correlates with each x in $\mathbf{X}^1$ the set of all y in $\mathbf{Y}^1$ such that $\mathbf{R}^2(x, y)$. Now consider the following sentence Cont:

$$(\exists \mathbf{X}^1)(\exists \mathbf{Y}^1)(\mathrm{Den}(\mathbf{X}^1) \wedge (\forall y)\mathbf{Y}^1(y) \wedge \mathscr{P}(\mathbf{Y}^1, \mathbf{X}^1))$$

†For a definition of *elementary submodel*, see Section 2.13.

‡The axioms are those for an ordered field (see p.99) plus a second-order completeness axiom. The latter can be taken to be the assertion that every nonempty subset that is bounded above has a least upper bound (or, equivalently, that no Dedekind cut is a gap). For a proof of categoricity, see Mendelson [1973], Section 5.4.

Then Cont is true in a standard interpretation if and only if the domain of the interpretation has the power of the continuum, since the power set of a denumerable set has the power of the continuum. See Shapiro (1991, Section 5.1.2) and Garland (1974) for more information about the definability of cardinal numbers in second-order logic.

Exercises

A.7 Show that a sentence of pure second-order logic is true in a standard interpretation $\mathscr{M}$ if and only if it is true in any other standard interpretation whose domain has the same cardinal number as that of $\mathscr{M}$.

A.8 (a) Show that there is a formula Cont ($\mathbf{X}^1$) of pure second-order logic that is satisfied by an assignment in an interpretation if and only if the set assigned to $\mathbf{X}^1$ has the power of the continuum.

(b) Find a sentence CH of pure second-order logic that is standardly valid if and only if the continuum hypothesis is true.†

HENKIN SEMANTICS FOR L2C

In light of the fact that completeness, compactness and the Skolem–Löwenheim theorems do not hold in second-order logic, it is of some interest that there is a modification of the semantics for second-order logic that removes those drawbacks and restores a completeness property. The fundamental ideas sketched below are due to Henkin (1950).

Start with a first-order interpretation with domain D. For each positive integer n, choose a fixed collection $\mathscr{D}(n)$ of n-ary relations on D, and a fixed collection $\mathscr{F}(n)$ of n-ary operations on D. Instead of $\sum_2$, we now use the set $\sum_2^{\mathrm{H}}$ of assignments s in $\sum_2$ such that, for each predicate variable $\mathbf{R}^n$, $s(\mathbf{R}^n)$ is in $\mathscr{D}(n)$ and, for each function variable $\mathbf{g}^n$, $s(\mathbf{g}^n)$ is in $\mathscr{F}(n)$. The definitions of satisfaction and truth are the same as for standard semantics, except that $\sum_2$ is replaced by $\sum_2^{\mathrm{H}}$. Such an interpretation will be called a *Henkin interpretation*. Using a Henkin interpretation amounts to restricting the ranges of the predicate and function variables. For example, the range of a predicate variable $\mathbf{R}^1$ need not be the entire power set $\mathscr{P}(D)$ of the domain D. In order for a Henkin interpretation $\mathscr{H}$ to serve as an adequate semantic framework, we must require that all instances of the comprehension schema and the function definition schema are true in $\mathscr{H}$. A Henkin interpretation

†We take as the continuum hypothesis the assertion that every subset of the set of real numbers is either finite or denumerable or is equinumerous with the set of all real numbers.

for which this condition is met will be called a *general model.* A formula that is true in all general models will be said to be *generally valid,* and a formula that is satisfied by some assignment in some general model will be said to be *generally satisfiable.* We say that $\mathscr{B}$ *generally implies* $\mathscr{C}$ if $\mathscr{B} \Rightarrow \mathscr{C}$ is generally valid, and that $\mathscr{B}$ is *generally equivalent* to $\mathscr{C}$ if $\mathscr{B} \Leftrightarrow \mathscr{C}$ is generally valid.

A standard interpretation on a domain D determines a corresponding general model in which $\mathscr{D}(n)$ is the set of *all* n-ary relations on D and $\mathscr{F}(n)$ is the set of *all* n-ary operations on D. Such a general model is called a *full* general model. Standard satisfaction and truth are equivalent to Henkin satisfaction and truth for the corresponding full general model. Hence, the following statements are obvious.

PROPOSITION A.9

(a) Every generally valid formula is also standardly valid.
(b) Every standardly satisfiable formula is generally satisfiable.

We also have the following strengthening of Proposition A1.

PROPOSITION A.10

Every theorem of PC2 is generally valid.

Proof

The general validity of (Comp) and (FunDef) follows from the definition of a general model. The proofs for the other logical axioms are similar to those in the first-order case, as is the verification that general validity is preserved by the rules of inference.

PROPOSITION A.11 (GENERAL SECOND-ORDER COMPLETENESS)

The theorems of PC2 coincide with the generally valid formulas of L2C.

Proof

Let $\mathscr{B}$ be a generally valid formula of L2C. We must show that $\mathscr{B}$ is a theorem of PC2. (It suffices to consider only closed formulas.) Assume, for the sake of contradiction, that $\mathscr{B}$ is not a theorem of PC2. Then, by the deduction theorem, the theory PC2 + $\{\neg\mathscr{B}\}$ is consistent. If we could prove

that any consistent extension of PC2 has a general model, then it would follow that PC2 + $\{\neg\mathscr{B}\}$ has a general model, contradicting our hypothesis that $\mathscr{B}$ is generally valid. Hence, it suffices to establish the following result.

HENKIN'S LEMMA

Every consistent extension $\mathscr{T}$ of PC2 has a general model.

Proof

The strategy is the same as in Henkin's proof of the fact that every consistent first-order theory has a model. One first adds enough new individual constants, function letters and predicate letters to provide 'witnesses' for all existential sentences. For example, for each sentence $(\exists x)\mathscr{C}(x)$ there will be a new individual constant b such that $(\exists x)\mathscr{C}(x) \Rightarrow \mathscr{C}(b)$ can be consistently added to the theory. (See Lemma 2.15 for the basic technique.) The same thing is done for existential quantifiers $(\exists \mathbf{g}^n)$ and $(\exists \mathbf{R}^n)$. Let $\mathscr{T}^*$ be the consistent extension of $\mathscr{T}$ obtained by adding all such conditionals as axioms. Then, by the method of Lindenbaum's lemma (Lemma 2.14), we inductively extend $\mathscr{T}^*$ to a maximal consistent theory $\mathscr{T}^\#$. A general model $\mathscr{M}$ of $\mathscr{T}$ can be extracted from $\mathscr{T}^\#$. The domain consists of the constant terms of $\mathscr{T}^\#$. The range of the predicate variables consists of the relations determined by the predicate letters of $\mathscr{T}^\#$. A predicate letter B determines the relation $B^\#$ such that $B^\#\langle t\rangle_n$ holds in $\mathscr{M}$ if and only if $B^\#\langle t\rangle_n$ is a theorem of $\mathscr{T}^\#$. The range of the function variables consists of the operations determined by the function letters of $\mathscr{T}^\#$. If f is a function letter of $\mathscr{T}^\#$, define an operation $f^\#$ by letting $f^\#(\langle t\rangle_n) = f(\langle t\rangle_n)$. A proof by induction shows that, for every sentence $\mathscr{C}$, $\mathscr{C}$ is true in $\mathscr{M}$ if and only if $\mathscr{C}$ is a theorem of $\mathscr{T}^\#$. In particular, all theorems of $\mathscr{T}$ are true in $\mathscr{M}$.

The compactness property and the Skolem–Löwenheim theorems also hold for general models. See Manzano (1996, Chapter IV), or Shapiro (1991) for detailed discussions.†

COROLLARY A.12

There are standardly valid formulas that are not generally valid.

†Lindström (1969) has shown that, in a certain very precise sense, first-order logic is the strongest logic that satisfies the countable compactness and Skolem–Löwenheim theorems. So, general models really are disguised first-order models.

Proof

By Corollary A.7, there is no axiomatic formal system whose theorems are the standardly valid formulas of L2$\mathscr{A}$. By Proposition A.11, the generally valid formulas of L2$\mathscr{A}$ are the theorems of the second-order theory P$\mathscr{A}$2. Hence, the set of standardly valid formulas of L2$\mathscr{A}$ is different from the set of generally valid formulas of L2$\mathscr{A}$. Since all generally valid formulas are standardly valid, there must be some standardly valid formula that is not generally valid.

We can exhibit an explicit sentence that is standardly valid but not generally valid. The Gödel–Rosser incompleteness theorem (Proposition 3.38) can be proved for the second-order theory AR2. Let $\mathscr{R}$ be Rosser's undecidable sentence for AR2.† If AR2 is consistent, $\mathscr{R}$ is true in the standard model of arithmetic. (Recall that $\mathscr{R}$ asserts that, for any proof in AR2 of $\mathscr{R}$, there is a proof in AR2, with a smaller Gödel number, of $\neg\mathscr{R}$. If AR2 is consistent, $\mathscr{R}$ is undecidable in AR2 and, therefore, there is no proof in AR2 of $\mathscr{R}$, which makes $\mathscr{R}$ trivially true.) Hence, AR2 $\Rightarrow \mathscr{R}$ is standardly valid, by Proposition A.4. However, AR2 $\Rightarrow \mathscr{R}$ is not generally valid. For, if AR2 $\Rightarrow \mathscr{R}$ were generally valid, it would be provable in P$\mathscr{A}$2, by Proposition A.11. Hence, $\mathscr{R}$ would be provable in AR2, contradicting the fact that it is an undecidable sentence of AR2.

Exercise

A.9 (a) Show that the second-order theory AR2 is recursively undecidable.

(b) Show that the pure second-order predicate calculus P$\mathscr{A}$2 is recursively undecidable.‡

It appears that second-order and higher-order logics were the implicitly understood logics of mathematics until the 1920s. The axiomatic characterization of the natural numbers by Dedekind and Peano, the axiomatic characterization of the real numbers as a complete ordered field by Hilbert in 1900, and Hilbert's axiomatization of Euclidean geometry in 1902 (in the French translation of his original 1899 book) all presupposed a second-order logic in order to obtain the desired categoricity. The distinction between first-order and second-order languages was made by Löwenheim (1915) and by Hilbert in unpublished 1917 lectures, and was crystal-clear in

†We must assume that AR is consistent.

‡The pure second-order monadic predicate logic MP2 (in which there are no nonlogical constants and no function variables, and all second-order predicate variables are monadic) is recursively decidable. See Ackermann (1954) for a proof. The earliest proof was found by Löwenheim (1915), and simpler proofs were given by Skolem (1919) and Behmann (1922).

Hilbert and Ackermann's (1950),[†] where the problem was posed about the completeness of their axiom system for first-order logic. The positive solution to this problem presented in Gödel (1930), and the compactness and Skolem–Löwenheim theorems that followed therefrom, probably made the use of first-order logic more attractive. Another strong point favoring first-order logic was the fact that Skolem in 1922 constructed a first-order system for axiomatic set theory that overcame the imprecision in the Zermelo and Fraenkel systems.[‡] Skolem was always an advocate of first-order logic, perhaps because it yielded the relativity of mathematical notions that Skolem believed in. Philosophical support for first-order logic came from W.V. Quine, who championed the position that *logic* is first-order logic, and that second-order logic is just set theory in disguise.

The rich lodes of first-order model theory and proof theory kept logicians busy and satisfied for over a half-century, but recent years have seen a revival of interest in higher-order logic and other alternatives to first-order logic, and the papers in the book *Model-Theoretic Logics* (edited by Barwise and Feferman (1985)) offer a picture of these new developments.[§] Barwise (1985) lays down the challenge to the old first-order orthodoxy, and Shapiro (1991) and Corcoran (1987) provide philosophical, historical and technical support for higher-order logic. Of course, we need not choose between first-order and higher-order logic; there is plenty of room for both.

[†]Hilbert and Ackermann (1950) is a translation of the second (1938) edition of a book which was first published in 1928 as *Grundzüge der theoretischen Logik.*

[‡]See Moore (1988) and Shapiro (1991) for more about the history of first-order logic. Shapiro (1991) is the most reliable and thorough study of the controversies involving first-order and second-order logic.

[§]Van Benthem and Doets (1983) also provides a high-level survey of second-order logic and its ramifications.

Answers to Selected Exercises

CHAPTER 1

1.1

A	B	
T	T	F
F	T	T
T	F	T
F	F	F

1.2

A	B	$\neg A$	$A \Rightarrow B$	$(A \Rightarrow B) \vee \neg A$
T	T	F	T	T
F	T	T	T	T
T	F	F	F	F
F	F	T	T	T

1.3

$((A$	$\Rightarrow$	$B)$	$\wedge$	$A)$
T	T	T	T	T
F	T	T	F	F
T	F	F	F	T
F	T	F	F	F

1.4 (a) $((A \Rightarrow (\neg B)) \wedge ((\neg A) \Rightarrow (\neg B)))$

(c) $(A \Rightarrow B)$, A: x is prime, B: x is odd.

(d) $(A \Rightarrow B)$, A: the sequence s converges, B: the sequence s is bounded.

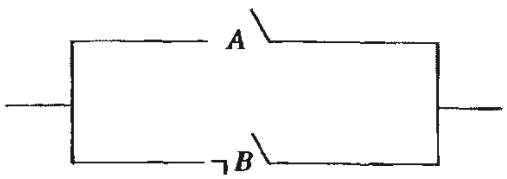
(e) $(A \Leftrightarrow (B \wedge (C \wedge D)))$
A: the sheikh is happy,
B: the sheikh has wine,
C: the sheikh has women,
D: the sheikh has song.

(f) $(A \Rightarrow B)$, A: Fiorello goes to the movies.

(i) $((\neg A) \Rightarrow B)$, A: Kasparov wins today, B: Karpov will win the tournament.

1.5 (c), (d), (f), (g), (i), (j) are tautologies.

1.6 (a), (b), (d), (e), (f) are logically equivalent pairs.
1.11 All except (i).
1.13 Only (c) and (e).
1.15 (a) $(B \Rightarrow \neg A) \wedge C$ (e) $A \Leftrightarrow B \Leftrightarrow \neg(C \vee D)$
(c) Drop all parentheses. (g) $\neg(\neg\neg(B \vee C) \Leftrightarrow (B \Leftrightarrow C))$
1.16 (a) $(C \vee ((\neg A) \wedge B))$ (c) $((C \Rightarrow ((\neg((A \vee B) \Rightarrow C)) \wedge A)) \Leftrightarrow B)$
1.17 (a) $(((\neg(\neg A)) \Leftrightarrow A) \Leftrightarrow (B \vee C))$ (d) and (f) are the only ones that are not abbreviations of statement forms.
1.18 (a) $\vee \Rightarrow C\neg AB$ and $\vee C \Rightarrow \wedge B\neg DC$
(c) (a) $\wedge \Rightarrow B\neg AC$ (b) $\vee A \vee BC$
(d) (i) is not. (ii) $(A \Rightarrow B) \Rightarrow ((B \Rightarrow C) \Rightarrow (\neg A \Rightarrow C))$
1.19 (f) is contradictory, and (a), (d), (e), (g)–(j) are tautologies.
1.20 (b)–(d) are false.
1.21 (a) T (b) T (c) indeterminate
1.22 (a) A is T, B is F, and $\neg A \vee (A \Rightarrow B)$ is F.
(c) A is T, C is T, B is T.
1.29 (c) (i) $A \wedge ((B \wedge C) \vee (\neg B \wedge \neg C))$ (ii) $A \wedge B \wedge \neg C$
(iii) $\neg A \vee (\neg B \wedge C)$
1.30 (a) If $\mathscr{B}$ is a tautology, the result of replacing all statement letters by their negations is a tautology. If we then move all negation signs outward by using Exercise 1.27 (k) and (l), the resulting tautology is $\neg\mathscr{B}'$. Conversely, if $\neg\mathscr{B}'$ is a tautology, let $\mathscr{C}$ be $\neg\mathscr{B}'$. By the first part, $\neg\mathscr{C}'$ is a tautology. But $\neg\mathscr{C}'$ is $\neg\neg\mathscr{B}$.
(c) $(\neg A \wedge \neg B \wedge \neg C) \vee (A \wedge B \wedge \neg D)$
1.32 (a) For figure 1.4:



1.33 (a), (d) and (h) are not correct.
1.34 (a) Satisfiable: Let A, B, and C be F, and let D be T.
1.36 For f,

$$(A \wedge B \wedge C) \vee (\neg A \wedge B \wedge C) \vee (A \wedge \neg B \wedge C) \vee (\neg A \wedge \neg B \wedge \neg C)$$

1.37 For $\Rightarrow$ and $\vee$, notice that any statement form built up using $\Rightarrow$ and $\vee$ will always take the value T when the statement letters in it are T. In the case of $\neg$ and $\Leftrightarrow$, using only the statement letters A and B, find all the truth functions of two variables that can be generated by applying $\neg$ and $\Leftrightarrow$ any number of times.
1.40 (a) $2^4 = 16$ (b) 2^{2^n}

1.41 $h(C,C,C) = \neg C$ and $h(B,B,\neg C)$ is $B \Rightarrow C$.

1.42 (b) For $\neg(A \Rightarrow B) \vee (\neg A \wedge C)$, a disjunctive normal form is $(A \wedge \neg B) \vee (\neg A \wedge C)$, and a conjunctive normal form is $(A \vee C) \wedge (\neg B \vee \neg A) \wedge (\neg B \vee C)$.
(c) (i) For $(A \wedge B) \vee \neg A$, a full dnf is $(A \wedge B) \vee (\neg A \wedge B) \vee (\neg A \wedge \neg B)$, and a full cnf is $B \vee \neg A$.

1.43 (b) (i) Yes. A: T, B: T, C: F (ii) Yes. A: T, B: F, C: T

1.45 (b) A conjunction $\mathscr{E}$ of the form $B_1^* \wedge \ldots \wedge B_n^*$, where each B_i^* is either B_i or $\neg B_i$, is said to be *eligible* if some assignment of truth values to the statement letters of $\mathscr{B}$ that makes $\mathscr{B}$ true also makes $\mathscr{E}$ true. Let $\mathscr{C}$ be the disjunction of all eligible conjunctions.

1.47 (b)

1.	$\mathscr{C} \Rightarrow \mathscr{D}$	Hypothesis
2.	$\mathscr{B} \Rightarrow \mathscr{C}$	Hypothesis
3.	$(\mathscr{B} \Rightarrow (\mathscr{C} \Rightarrow \mathscr{D})) \Rightarrow ((\mathscr{B} \Rightarrow \mathscr{C}) \Rightarrow (\mathscr{B} \Rightarrow \mathscr{D}))$	Axiom (A2)
4.	$(\mathscr{C} \Rightarrow \mathscr{D}) \Rightarrow (\mathscr{B} \Rightarrow (\mathscr{C} \Rightarrow \mathscr{D}))$	Axiom (A1)
5.	$\mathscr{B} \Rightarrow (\mathscr{C} \Rightarrow \mathscr{D})$	1, 4, MP
6.	$(\mathscr{B} \Rightarrow \mathscr{C}) \Rightarrow (\mathscr{B} \Rightarrow \mathscr{D})$	3, 5, MP
7.	$\mathscr{B} \Rightarrow \mathscr{D}$	2, 6, MP

1.48 (a)

1.	$\mathscr{B} \Rightarrow \neg\neg B$	Lemma 1.11(b)
2.	$\neg\neg AB \Rightarrow (\neg B \Rightarrow \mathscr{C})$	Lemma 1.11(c)
3.	$\mathscr{B} \Rightarrow (\neg\mathscr{B} \Rightarrow \mathscr{C})$	1, 2, Corollary 1.10(a)
4.	$\mathscr{B} \Rightarrow (\mathscr{B} \vee \mathscr{C})$	3, Abbreviation

(c)

1.	$\neg\mathscr{C} \Rightarrow \mathscr{B}$	Hypothesis
2.	$(\neg\mathscr{C} \Rightarrow \mathscr{B}) \Rightarrow (\neg\mathscr{B} \Rightarrow \neg\neg\mathscr{C})$	Lemma 1.11(e)
3.	$\neg\mathscr{B} \Rightarrow \neg\neg\mathscr{C}$	1, 2, MP
4.	$\neg\neg\mathscr{C} \Rightarrow \mathscr{C}$	Lemma 1.11(a)
5.	$\neg\mathscr{B} \Rightarrow \mathscr{C}$	3, 4, Corollary 1.10(a)
6.	$\neg\mathscr{C} \Rightarrow \mathscr{B} \vdash \neg\mathscr{B} \Rightarrow \mathscr{C}$	1–5
7.	$\vdash (\neg\mathscr{C} \Rightarrow \mathscr{B}) \Rightarrow (\neg\mathscr{B} \Rightarrow \mathscr{C})$	6, deduction theorem
8.	$\vdash (\mathscr{C} \vee \mathscr{B}) \Rightarrow (\mathscr{B} \vee \mathscr{C})$	7, abbreviation

1.50 Take any assignment of truth values to the statement letters of $\mathscr{B}$ that makes $\mathscr{B}$ false. Replace in $\mathscr{B}$ each letter having the value T by $A_1 \vee \neg A_1$, and each letter having the value F by $A_1 \wedge \neg A_1$. Call the resulting statement form $\mathscr{C}$. Thus, $\mathscr{C}$ is an axiom of L*, and , therefore, $\vdash_{L^*} \mathscr{C}$. Observe that $\mathscr{C}$ always has the value F for any truth assignment. Hence, $\neg\mathscr{C}$ is a tautology. So $\vdash_L \neg\mathscr{C}$ and, therefore, $\vdash_{L^*} \neg\mathscr{C}$.

1.51 (Deborah Moll) Use two truth values. Let $\Rightarrow$ have its usual table and let $\neg$ be interpreted as the constant function F. When B is F, $(\neg B \Rightarrow \neg A) \Rightarrow ((\neg B \Rightarrow A) \Rightarrow B)$ is F.

1.52 The theorems of P are the same as the axioms. Assume that P is suitable for some n-valued logic. Then, for all values k, $k * k$ will be a designated value. Consider the sequence of formulas $\mathscr{B}_0 = A$, $\mathscr{B}_{j+1} = A * \mathscr{B}_j$. Since there are n^n possible truth functions of one variable, among $\mathscr{B}_0, \ldots, \mathscr{B}_{n^n}$ there must be two different formulas $\mathscr{B}_j$ and $\mathscr{B}_k$ that determine

the same truth function. Hence, $\mathscr{B}_j * \mathscr{B}_k$ will be an exceptional formula that is not a theorem.

1.53 Take as axioms all exceptional formulas, and the identity function as the only rule of inference.

CHAPTER 2

2.1 (a) $((\forall x_1)(A_1^1(x_1) \wedge (\neg A_1^1(x_2))))$ (b) $(((\forall x_2)A_1^1(x_2)) \Leftrightarrow A_1^1(x_2))$
(d) $(((\forall x_1)((\forall x_3)((\forall x_4)A_1^1(x_1)))) \Rightarrow (A_1^1(x_2) \wedge (\neg A_1^1(x_1))))$

2.2 (a) $((\forall x_1)(A_1^1(x_1) \Rightarrow A_1^1(x_1))) \vee (\exists x_1)A_1^1(x_1)$

2.3 (a) The only free occurrence of a variable is that of x_2.
(b) The first occurrence of x_3 is free, as is the last occurrence of x_2.

2.6 Yes, in parts (a), (c) and (e)

2.8 (a) $(\forall x)(P(x) \Rightarrow L(x))$
(b) $(\forall x)(P(x) \Rightarrow \neg H(x))$ or $\neg(\exists x)(P(x) \wedge H(x))$
(c) $\neg(\forall x)(B(x) \Rightarrow F(x))$
(d) $(\forall x)(B(x) \Rightarrow \neg F(x))$ (e) $T(x) \Rightarrow I(x)$
(f) $(\forall x)(\forall y)(S(x) \wedge D(x,y) \Rightarrow J(y))$
(j) $(\forall x)(\neg H(x,x) \Rightarrow H(j,x))$ or $(\forall x)(P(x) \wedge \neg H(x,x) \Rightarrow H(j,x))$

(In the second wf, we have specified that John hates those *persons* who do not hate themselves, where $P(x)$ means x *is a person.*)

2.9 (a) All bachelors are unhappy. (c) There is no greatest integer.

2.10 (a) (i) is satisfied by all pairs $\langle x_1, x_2 \rangle$ of positive integers such that $x_1 \cdot x_2 \geq 2$.
(ii) is satisfied by all pairs $\langle x_1, x_2 \rangle$ of positive integers such that either $x_1 < x_2$ (when the antecedent is false) or $x_1 = x_2$ (when the antecedent and consequent are both true).
(iii) is true.

2.11 (a) Between any two real numbers there is a rational number.

2.12 (I) A sequence s satisfies $\neg\mathscr{B}$ if and only if s does not satisfy $\mathscr{B}$. Hence, all sequences satisfy $\neg\mathscr{B}$ if and only if no sequence satisfies $\mathscr{B}$; that is, $\neg\mathscr{B}$ is true if and only if $\mathscr{B}$ is false.

(II) There is at least one sequence s in $\sum$. If s satisfies $\mathscr{B}$, $\mathscr{B}$ cannot be false for M. If s does not satisfy $\mathscr{B}$, $\mathscr{B}$ cannot be true for M.

(III) If a sequence s satisfies both $\mathscr{B}$ and $\mathscr{B} \Rightarrow \mathscr{C}$, then s satisfies $\mathscr{C}$ by condition 3 of the definition.

(V) (a) s satisfies $\mathscr{B} \wedge \mathscr{C}$ if and only if s satisfies $\neg(\mathscr{B} \Rightarrow \neg\mathscr{C})$
if and only if s does not satisfy $\mathscr{B} \Rightarrow \neg\mathscr{C}$
if and only if s satisfies $\mathscr{B}$ but not $\neg\mathscr{C}$
if and only if s satisfies $\mathscr{B}$ and s satisfies $\mathscr{C}$

(VI)(a) Assume $\models_M \mathscr{B}$. Then every sequence satisfies $\mathscr{B}$. In particular, every sequence that differs from a sequence s in at most the ith

place satisfies $\mathscr{B}$. So, every sequence satisfies $(\forall x_i)\mathscr{B}$; that is, $\models_{\mathrm{M}} (\forall x_i)\mathscr{B}$.

(b) Assume $\models_{\mathrm{M}} (\forall x_i)\mathscr{B}$. If s is a sequence, then any sequence that differs from s in at most the ith place satisfies $\mathscr{B}$, and, in particular, s satisfies $\mathscr{B}$. Then every sequence satisfies $\mathscr{B}$; that is, $\models_{\mathrm{M}} \mathscr{B}$.

(VIII) *Lemma.* If all the variables in a term t occur in the list $x_{i_1}, \ldots, x_{i_k}$ ($k \geq 0$; when $k = 0$, t has no variables), and if the sequences s and s' have the same components in the i_1th, $\ldots$, i_kth places, then $s^*(t) = (s')^*(t)$.

Proof. Induction on the number m of function letter in t. Assume the result holds for all integers less than m.

Case 1. t is an individual constant a_p. Then $s^*(a_p) = (a_p)^{\mathrm{M}} = (s')^*(a_p)$.

Case 2. t is a variable x_{i_j}. Then $s^*(x_{i_j}) = s_{i_j} = s'_{i_j} = (s')^*(x_{i_j})$.

Case 3. t is of the form $f_j^n(t_1, \ldots, t_n)$. For $q \leqslant n$, each t_q has fewer than m function letters and all its variables occur among $x_{i_1}, \ldots, x_{i_k}$. By inductive hypothesis, $s^*(t_q) = (s')^*(t_q)$. Then $s^*(f_j^n(t_1, \ldots, t_n)) = (f_j^n)^{\mathrm{M}}(s^*(t_1), \ldots, s^*(t_n)) = (f_j^n)^{\mathrm{M}}((s')^*(t_1), \ldots, (s')^*(t_n)) = (s')^*(f_j^n(t_1, \ldots, t_n))$.

Proof of (VIII). Induction on the number r of connectives and quantifiers in $\mathscr{B}$. Assume the result holds for all $q < r$.

Case 1. $\mathscr{B}$ is of the form $A_j^n(t_1, \ldots, t_n)$; that is, $r = 0$. All the variables of each t_i occur among $x_{i_1}, \ldots, x_{i_k}$. Hence, by the lemma, $s^*(t_i) = (s')^*(t_i)$. But s satisfies $A_j^n(t_1, \ldots, t_n)$ if and only if $\langle s^*(t_1), \ldots, s^*(t_n)\rangle$ is in $(A_j^n)^{\mathrm{M}}$ – that is, if and only if $\langle (s')^*(t_1), \ldots, (s')^*(t_n)\rangle$ is in $(A_j^n)^{\mathrm{M}}$, which is equivalent to s' satisfying $A_j^n(t_1, \ldots, t_n)$.

Case 2. $\mathscr{B}$ is of the form $\neg\mathscr{C}$.

Case 3. $\mathscr{B}$ is of the form $\mathscr{C} \Rightarrow \mathscr{D}$. Both cases 2 and 3 are easy.

Case 4. $\mathscr{B}$ is of the form $(\forall x_j)\mathscr{C}$. The free variables of $\mathscr{C}$ occur among $x_{i_1}, \ldots, x_{i_k}$ and x_j. Assume s satisfies $\mathscr{B}$. Then every sequence that differs from s in at most the jth place satisfies $\mathscr{C}$. Let $s^{\#}$ be any sequence that differs from s' in at most the jth place. Let s^b be a sequence that has the same components as s in all but the jth place, where it has the same component as $s^{\#}$. Hence, s^b satisfies $\mathscr{C}$. Since s^b and $s^{\#}$ agree in the i_1th,..., i_kth and jth places, it follows by inductive hypothesis that s^b satisfies $\mathscr{C}$ if and only if $s^{\#}$ satisfies $\mathscr{C}$. Hence, $s^{\#}$ satisfies $\mathscr{C}$. Thus, s' satisfies $\mathscr{B}$. By symmetry, the converse also holds.

(IX) Assume $\mathscr{B}$ is closed. By (VIII), for any sequence s and s', s satisfies $\mathscr{B}$ if and only if s' satisfies $\mathscr{B}$. If $\neg\mathscr{B}$ is not true for M, some sequence s' does not satisfy $\neg\mathscr{B}$; that is, s' satisfies $\mathscr{B}$. Hence, every sequence s satisfies $\mathscr{B}$; that is, $\models_{\mathrm{M}} \mathscr{B}$.

(X) *Proof of Lemma 1:* induction on the number m of function letters in t.

Case 1. t is a_j. Then t' is a_j. Hence,

$$s^*(t') = s^*(a_j) = (a_j)^{\mathrm{M}} = (s')^*(a_j) = (s')^*(t)$$

Case 2. t is x_j, where $j \neq i$. Then t' is x_j. By the lemma of (VIII), $s^*(t') = (s')^*(t)$, since s and s' have the same component in the jth place.

Case 3. t is x_i. Then t' is u. Hence, $s^*(t') = s^*(u)$, while $(s')^*(t) = (s')^*(x_i) = s'_i = s^*(u)$.

Case 4. t is of the form $f_j^n(t_1, \ldots, t_n)$. For $1 \leqslant q \leqslant n$, let t'_q result from t_q by the substitution of u for x_i. By inductive hypothesis, $s^*(t'_q) = (s')^*(t_q)$. But

$$s^*(t') = s^*(f_j^n(t'_1, \ldots, t'_n)) = (f_j^n)^{\mathrm{M}}(s^*(t'_1), \ldots, s^*(t'_n))$$
$$= (f_j^n)^{\mathrm{M}}((s')^*(t_1), \ldots, (s')^*(t_n)) = (s')^*(f_j^n(t_1, \ldots, t_n)) = (s')^*(t)$$

Proof of Lemma 2(a): induction on the number m of connectives and quantifiers in $\mathscr{B}(x_i)$.

Case 1. $m = 0$. Then $\mathscr{B}(x_i)$ is $A_j^n(t_1, \ldots, t_n)$. Let t'_q be the result of substituting t for all occurrences of x_i in t_q. Thus, $\mathscr{B}(t)$ is $A_j^n(t'_1, \ldots, t'_n)$. By Lemma 1, $s^*(t'_q) = (s')^*(t_q)$. Now, s satisfies $\mathscr{B}(t)$ if and only if $\langle s^*(t'_1), \ldots, s^*(t'_n)\rangle$ belongs to $(A_j^n)^{\mathrm{M}}$, which is equivalent to $\langle (s')^*(t_1), \ldots, (s')^*(t_n)\rangle$ belonging to $(A_j^n)^{\mathrm{M}}$ – that is, to s' satisfying $\mathscr{B}(x_i)$.

Case 2. $\mathscr{B}(x_i)$ is $\neg\mathscr{C}(x_i)$; this is straightforward.

Case 3. $\mathscr{B}(x_i)$ is $\mathscr{C}(x_i) \Longrightarrow \mathscr{D}(x_i)$; this is straightforward.

Case 4. $\mathscr{B}(x_i)$ is $(\forall x_j)\mathscr{B}(x_i)$.

Case 4a. x_j is x_i. Then x_i is not free in $\mathscr{B}(x_i)$, and $\mathscr{B}(t)$ is $\mathscr{B}(x_i)$. Since x_i is not free in $\mathscr{B}(x_i)$, it follows by (VIII) that s satisfies $\mathscr{B}(t)$ if and only if s' satisfies $\mathscr{B}(x_i)$.

Case 4b. x_j is different from x_i. Since t is free for x_i in $\mathscr{B}(x_i)$, t is also free for x_i in $\mathscr{C}(x_i)$.

Assume s satisfies $(\forall x_j)\mathscr{C}(t)$. We must show that s' satisfies $(\forall x_j)\mathscr{C}(x_i)$. Let $s^{\#}$ differ from s' in at most the jth place. It suffices to show that $s^{\#}$ satisfies $\mathscr{C}(x_i)$. Let s^b be the same as $s^{\#}$ except that it has the same ith component as s. Hence, s^b is the same as s except in its jth component. Since s satisfies $(\forall x_j)\mathscr{C}(t)$, s^b satisfies $\mathscr{C}(t)$. Now, since t is free for x_i in $(\forall x_j)\mathscr{C}(x_i)$, t does not contain x_j. (The other possibility, that x_i is not free in $\mathscr{C}(x_i)$, is handled as in case 4a.) Hence, by the lemma of (VIII), $(s^b)^*(t) = s^*(t)$. Hence, by the inductive hypothesis and the fact that $s^{\#}$ is obtained from s^b by substituting $(s^b)^*(t)$ for the ith component of s^b, it follows that $s^{\#}$ satisfies $\mathscr{C}(x_i)$, if and only if s^b satisfies $\mathscr{C}(t)$. Since s^b satisfies $\mathscr{C}(t)$, $s^{\#}$ satisfies $\mathscr{C}(x_i)$.

Conversely, assume s' satisfies $(\forall x_j)\mathscr{C}(x_i)$. Let s^b differ from s in at most the jth place. Let $s^\#$ be the same as s' except in the jth place, where it is the same as s^b. Then $s^\#$ satisfies $\mathscr{C}(x_i)$. As above, $s^*(t) = (s^b)^*(t)$. Hence, by the inductive hypothesis, s^b satisfies $\mathscr{C}(t)$ if and only if $s^\#$ satisfies $\mathscr{C}(x_i)$. Since $s^\#$ satisfies $\mathscr{C}(x_i)$, s^b satisfies $\mathscr{C}(t)$. Therefore, s satisfies $(\forall x_j)\mathscr{C}(t)$.

Proof of Lemma 2(b). Assume s satisfies $(\forall x_i)\mathscr{B}(x_i)$. We must show that s satisfies $\mathscr{B}(t)$. Let s' arise from s by substituting $s^*(t)$ for the ith component of s. Since s satisfies $(\forall x_i)\mathscr{B}(x_i)$ and s' differs from s in at most the ith place, s' satisfies $\mathscr{B}(x_i)$. By Lemma 2(a), s satisfies $\mathscr{B}(t)$.

2.13 Assume $\mathscr{B}$ is satisfied by a sequence s. Let s' be any sequence. By (VIII), s' also satisfies $\mathscr{B}$. Hence, $\mathscr{B}$ is satisfied by all sequences; that is, $\models_M \mathscr{B}$.

2.14 (a) x is a common divisor of y and z. (d) x_1 is a bachelor.

2.15 (a) (i) Every non-negative integer is even or odd. True.
(ii) If the product of two non-negative integers is zero, at least one of them is zero. True. (iii) 1 is even. False.

2.17 (a) Consider an interpretation with the set of integers as its domain. Let $A_1^1(x)$ mean that x is even and let $A_2^1(x)$ mean that x is odd. Then $(\forall x_1)A_1^1(x_1)$ is false, and so $(\forall x_1)A_1^1(x_1) \Longrightarrow (\forall x_1)A_2^1(x_1)$ is true. However, $(\forall x_1)(A_1^1(x_1) \Longrightarrow A_2^1(x_1))$ is false, since it asserts that all even integers are odd.

2.18 (a) $[(\forall x_i)\neg\mathscr{B}(x_i) \Longrightarrow \neg\mathscr{B}(t)] \Longrightarrow [\mathscr{B}(t) \Longrightarrow \neg(\forall x_i)\neg\mathscr{B}(x_i)]$ is logically valid because it is an instance of the tautology $(A \Longrightarrow \neg B) \Longrightarrow (B \Longrightarrow \neg A)$. By (X), $(\forall x_i)\neg\mathscr{B}(x_i) \Longrightarrow \neg\mathscr{B}(t)$ is logically valid. Hence, by (III), $\mathscr{B}(t) \Longrightarrow \neg(\forall x_i)\ \neg\mathscr{B}(x_i)$ is logically valid.

(b) Intuitive proof: If $\mathscr{B}$ is true for all x_i, then $\mathscr{B}$ is true for some x_i. Rigorous proof: Assume $(\forall x_i)\mathscr{B} \Longrightarrow (\exists x_i)\mathscr{B}$ is not logically valid. Then there is an interpretation M for which it is not true. Hence, there is a sequence s in $\sum$ such that s satisfies $(\forall x_i)\mathscr{B}$ and s does not satisfy $\neg(\forall x_i)\neg\mathscr{B}$. From the latter, s satisfies $(\forall x_i)\neg\mathscr{B}$. Since s satisfies $(\forall x_i)\mathscr{B}$, s satisfies $\mathscr{B}$, and, since s satisfies $(\forall x_i)\neg\mathscr{B}$, s satisfies $\neg\mathscr{B}$. But then s satisfies both $\mathscr{B}$ and $\neg\mathscr{B}$, which is impossible.

2.19 (b) Take the domain to be the set of integers and let $A_1^1(u)$ mean that u *is even*. A sequence s in which s_1 is even satisfies $A_1^1(x_1)$ but does not satisfy $(\forall x_1)A_1^1(x_1)$.

2.21 (a) Let the domain be the set of integers and let $A_1^2(x, y)$ mean that $x < y$. (b) Same interpretation as in (a).

2.22 (a) The premisses are (i) $(\forall x)(S(x) \Longrightarrow N(x))$ and (ii) $(\forall x)(V(x) \Longrightarrow \neg N(x))$, and the conclusion is $(\forall x)(V(x) \Longrightarrow \neg S(x))$. Intuitive proof: Assume $V(x)$. By (ii), $\neg N(x)$. By (i), $\neg S(x)$. Thus, $\neg S(x)$ follows from $V(x)$, and the conclusion holds. A more rigorous proof can be given along the lines of (I)–(XI), but a better proof will become available after the study of predicate calculi.

2.26 (a) $(\exists x)(\exists y)(A_1^1(x) \wedge \neg A_1^1(y))$

2.27 (a)

1.	$(\forall x)(\mathscr{B} \Longrightarrow \mathscr{C})$	Hyp
2.	$(\forall x)\mathscr{B}$	Hyp
3.	$(\forall x)(\mathscr{B} \Longrightarrow \mathscr{C}) \Longrightarrow (\mathscr{B} \Longrightarrow \mathscr{C})$	Axiom (A4)
4.	$\mathscr{B} \Longrightarrow \mathscr{C}$	1,3, MP
5.	$(\forall x)\mathscr{B} \Longrightarrow \mathscr{B}$	Axiom (A4)
6.	$\mathscr{B}$	2,5, MP
7.	$\mathscr{C}$	4,6, MP
8.	$(\forall x)\mathscr{C}$	7, Gen
9.	$(\forall x)(\mathscr{B} \Longrightarrow \mathscr{C}), (\forall x)\mathscr{B} \vdash (\forall x)\mathscr{C}$	1–8
10.	$(\forall x)(\mathscr{B} \Longrightarrow \mathscr{C}) \vdash (\forall x)\mathscr{B} \Longrightarrow (\forall x)\mathscr{C}$	1–9, Corollary 2.6
11.	$\vdash (\forall x)(\mathscr{B} \Longrightarrow \mathscr{C}) \Longrightarrow ((\forall x)\mathscr{B} \Longrightarrow (\forall x)\mathscr{C})$	1–10, Corollary 2.6

2.28 *Hint:* Assume $\vdash_K \mathscr{B}$. By induction on the number of steps in the proof of $\mathscr{B}$ in K, prove that, for any variables $y_1, \ldots, y_n (n \geq 0)$, $\vdash_{K^\#} (\forall y_1) \ldots (\forall y_n)\mathscr{B}$.

2.31 (a)

1.	$(\forall x)(\forall y)A_1^2(x,y)$	Hyp
2.	$(\forall y)A_1^2(x,y)$	1, Rule A4
3.	$A_1^2(x,x)$	2, Rule A4
4.	$(\forall x)A_1^2(x,x)$	3, Gen
5.	$(\forall x)(\forall y)A_1^2(x,y) \vdash (\forall x)A_1^2(x,x)$	1–4
6.	$\vdash (\forall x)(\forall y)A_1^2(x,y) \Longrightarrow (\forall x)A_1^2(x,x)$	1–5, Corollary 2.6

2.33 (a) $\vdash \neg(\forall x)\neg\neg\mathscr{B} \Longleftrightarrow \neg(\forall x)\neg\mathscr{B}$ by the replacement theorem and the fact that $\vdash \neg\neg\mathscr{B} \Longleftrightarrow \mathscr{B}$. Replace $\neg(\forall x)\neg\neg\mathscr{B}$ by its abbreviation $(\exists x)\neg\mathscr{B}$.

2.36 (b) $(\exists\varepsilon)(\varepsilon > 0 \wedge (\forall\delta)(\delta > 0 \Longrightarrow (\exists x)(|x - c| < \delta \wedge \neg|f(x) - f(c)| < \varepsilon)))$

2.37 (a) (i) Assume $\vdash \mathscr{B}$. By moving the negation step-by-step inward to the atomic wfs, show that $\vdash \neg\mathscr{B}^* \Longleftrightarrow \mathscr{C}$, where $\mathscr{C}$ is obtained from $\mathscr{B}$ by replacing all atomic wfs by their negations. But, from $\vdash \mathscr{B}$ it can be shown that $\vdash \mathscr{C}$. Hence, $\vdash \neg\mathscr{B}^*$. The converse follows by noting that $(\mathscr{B}^*)^*$ is $\mathscr{B}$.

(ii) Apply (i) to $\neg\mathscr{B} \vee \mathscr{C}$.

2.39

1.	$(\exists y)(\forall x)(A_1^2(x,y) \Longleftrightarrow \neg A_1^2(x,x))$	Hyp
2.	$(\forall x)(A_1^2(x,b) \Longleftrightarrow \neg A_1^2(x,x))$	1, Rule C
3.	$A_1^2(b,y) \Longleftrightarrow \neg A_1^2(b,b)$	2, Rule A4
4.	$\mathscr{C} \wedge \neg\mathscr{C}$	3, Tautology

($\mathscr{C}$ is any wf not containing b.) Use Proposition 2.10 and proof by contradiction.

2.46 (a) In step 4, b is not a *new* individual constant. It was already used in step 2.

2.49 Assume K is complete and let $\mathscr{B}$ and $\mathscr{C}$ be closed wfs of K such that $\vdash_K \mathscr{B} \vee \mathscr{C}$. Assume not-$\vdash_K \mathscr{B}$. Then, by completeness, $\vdash_K \neg\mathscr{B}$. Hence, by the tautology $\neg A \Longrightarrow ((A \vee \mathscr{B}) \Longrightarrow \mathscr{B})$, $\vdash_K \mathscr{B}$. Conversely, assume K is not complete. Then there is a sentence $\mathscr{B}$ of K such that not-$\vdash_K \mathscr{B}$ and not-$\vdash_K \neg\mathscr{B}$. However, $\vdash_K \mathscr{B} \vee \neg\mathscr{B}$.

2.50 See Tarski, Mostowski and Robinson (1953, pp. 15–16).

2.55 (b) It suffices to assume $\mathscr{B}$ is a closed wf. (Otherwise, look at the closure of $\mathscr{B}$.) We can effectively write all the interpretations on a finite domain $\{b_1, \ldots, b_k\}$. (We need only specify the interpretations of the symbols that occur in $\mathscr{B}$.) For every such interpretation, replace every wf $(\forall x)\mathscr{C}(x)$, where $\mathscr{C}(x)$ has no quantifiers, by $\mathscr{C}(b_1) \wedge \ldots \wedge \mathscr{C}(b_k)$, and continue until no quantifiers are left. One can then evaluate the truth of the resulting wf for the given interpretation.

2.59 Assume K is not finitely axiomatizable. Let the axioms of K_1 be $\mathscr{B}_1, \mathscr{B}_2, \ldots$, and let the axioms of K_2 be $\mathscr{C}_1, \mathscr{C}_2, \ldots$. Then $\{\mathscr{B}_1, \mathscr{C}_1, \mathscr{B}_2, \mathscr{C}_2, \ldots\}$ is consistent. (If not, some finite subset $\{\mathscr{B}_1, \mathscr{B}_2, \ldots, \mathscr{B}_k, \mathscr{C}_1, \ldots, \mathscr{C}_m\}$ is inconsistent. Since K_1 is not finitely axiomatizable, there is a theorem $\mathscr{B}$ of K_1 such that $\mathscr{B}_1, \mathscr{B}_2, \ldots, \mathscr{B}_k \vdash \mathscr{B}$ does not hold. Hence, the theory with axioms $\{\mathscr{B}_1, \mathscr{B}_2, \ldots, \mathscr{B}_k, \neg\mathscr{B}\}$ has a model M. Since $\vdash_K \mathscr{B}$, M must be a model of K_2, and, therefore, M is a model of $\{\mathscr{B}_1, \mathscr{B}_2, \ldots, \mathscr{B}_k, \mathscr{C}_1, \ldots, \mathscr{C}_m\}$, contradicting the inconsistency of this set of wfs.) Since $\{\mathscr{B}_1, \mathscr{C}_1, \mathscr{B}_2, \mathscr{C}_2, \ldots\}$ is consistent, it has a model, which must be a model of both K_1 and K_2.

2.60 *Hint*: Let the closures of the axioms of K be $\mathscr{B}_1, \mathscr{B}_2, \ldots$. Choose a subsequence $\mathscr{B}_{j_1}, \mathscr{B}_{j_2}, \ldots$ such that $\mathscr{B}_{j_{n+1}}$ is the first sentence (if any) after $\mathscr{B}_{j_n}$ that is not deducible from $\mathscr{B}_{j_1} \wedge \ldots \wedge \mathscr{B}_{j_n}$. Let $\mathscr{C}_k$ be $\mathscr{B}_{j_1} \wedge \mathscr{B}_{j_2} \wedge \cdots \wedge \mathscr{B}_{j_k}$. Then the $\mathscr{C}_k$s form an axiom set for the theorems of K such that $\vdash \mathscr{C}_{k+1} \Longrightarrow \mathscr{C}_k$ but not-$\vdash \mathscr{C}_k \Longrightarrow \mathscr{C}_{k+1}$. Then $\{\mathscr{C}_1, \mathscr{C}_1 \Longrightarrow \mathscr{C}_2, \mathscr{C}_2 \Longrightarrow \mathscr{C}_3, \ldots\}$ is an independent axiomatization of K.

2.61 Assume $\mathscr{B}$ is not logically valid. Then the closure $\mathscr{C}$ of $\mathscr{B}$ is not logically valid. Hence, the theory K with $\neg\mathscr{C}$ as its only proper axiom has a model. By the Skolem–Löwenheim theorem, K has a denumerable model and, by the lemma in the proof of Corollary 2.22, K has a model of cardinality $\mathfrak{m}$. Hence, $\mathscr{C}$ is false in this model and, therefore, $\mathscr{B}$ is not true in some model of cardinality $\mathfrak{m}$.

2.65 (c)

1.	$x = x$	Proposition 2.23(a)
2.	$(\exists y)x = y$	1, rule E4
3.	$(\forall x)(\exists y)x = y$	2, Gen

2.68 (a) The problem obviously reduces to the case of substitution for a single variable at a time: $\vdash x_1 = y_1 \Longrightarrow t(x_1) = t(y_1)$. From (A7), $\vdash x_1 = y_1 \Longrightarrow (t(x_1) = t(x_1) \Longrightarrow t(x_1) = t(y_1))$. By Proposition 2.23 (a), $\vdash t(x_1) = t(x_1)$. Hence, $\vdash x_1 = y_1 \Longrightarrow t(x_1) = t(y_1)$.

2.70 (a) By Exercise 2.65(c), $\vdash (\exists y)x = y$. By Proposition 2.23(b,c), $\vdash (\forall y)(\forall z)(x = y \wedge x = z \Longrightarrow y = z)$. Hence, $\vdash (\exists_1 y)x = y$. By Gen, $\vdash (\forall x)(\exists_1 y)x = y$.

2.71 (b) (i) Let $\wedge_{1 \leqslant i < j \leqslant n} x_i \neq x_j$ stand for the conjunction of all wfs of the form $x_i \neq x_j$, where $1 \leqslant i < j \leqslant n$. Let $\mathscr{B}_n$ be $(\exists x_1) \ldots (\exists x_n) \wedge_{1 \leqslant i < j \leqslant n} x_i \neq x_j$.

(ii) Assume there is a theory with axioms $\mathscr{A}_1, \ldots, \mathscr{A}_n$ that has the same theorems as K. Each of $\mathscr{A}_1, \ldots, \mathscr{A}_n$ is provable from K_1 plus a finite number of the wfs $\mathscr{B}_1, \mathscr{B}_2, \ldots$. Hence, K_1 plus a finite number of wfs $\mathscr{B}_{j_1}, \ldots, \mathscr{B}_{j_n}$ suffices to prove all theorems of K. We may assume $j_1 < \cdots < j_n$. Then an interpretation whose domain consists of j_n objects would be a model of K, contradicting the fact that $\mathscr{B}_{j_n+1}$ is an axiom of K.

2.74 For the independence of axioms (A1) – (A3), replace all $t = s$ by the statement form $A \Rightarrow A$; then erase all quantifiers, terms and associated commas and parentheses; axioms (A4) – (A6) go over into statement forms of the form $P \Rightarrow P$, and axiom (A7) into $(P \Rightarrow P) \Rightarrow (Q \Rightarrow Q)$. For the independence of axiom (A1), the following four-valued logic, due to Dr D.K. Roy, works, where 0 is the only designated value.

A	B	$A \Rightarrow B$	A	B	$A \Rightarrow B$	A	B	$A \Rightarrow B$	A	B	$A \Rightarrow B$	A	$\neg A$
0	0	0	1	0	0	2	0	0	3	0	0	0	1
0	1	1	1	1	0	2	1	0	3	1	1	1	0
0	2	1	1	2	0	2	2	0	3	2	1	2	0
0	3	1	1	3	0	2	3	0	3	3	0	3	0

When A and B take the values 3 and 0, respectively, axiom (A1) takes the value 1. For the independence of axiom (A2), Dr Roy devised the following four-valued logic, where 0 is the only designated value.

A	B	$A \Rightarrow B$	A	B	$A \Rightarrow B$	A	B	$A \Rightarrow B$	A	B	$A \Rightarrow B$	A	$\neg A$
0	0	0	1	0	0	2	0	0	3	0	0	0	1
0	1	1	1	1	0	2	1	0	3	1	0	1	0
0	2	1	1	2	0	2	2	0	3	2	1	2	0
0	3	1	1	3	0	2	3	0	3	3	0	3	0

If A, B and C take the values 3, 0 and 2 respectively, then axiom (A2) is 1. For the independence of axiom (A3), the proof on page 44 works. For axiom (A4), replace all universal quantifiers by existential quantifiers. For axiom (A5), change all terms t to x_1 and replace all universal quantifiers by $(\forall x_1)$. For axiom (A6), replace all wfs $t = s$ by the negation of some fixed theorem. For axiom (A7), consider an interpretation in which the interpretation of $=$ is a reflexive non-symmetric relation.

2.83 (a) $(\forall x)(\exists y)((\exists z)(\mathscr{B}(z,x,y,\ldots,y) \wedge A_1^3(x,y,z)) \Rightarrow (\exists z)(\mathscr{B}(z,y,x,\ldots,x) \wedge z = x))$

2.84 (a) $(\exists z)(\forall w)(\exists x)([A_1^1(x) \Rightarrow A_1^2(x,y)] \Rightarrow [A_1^1(w) \Rightarrow A_1^2(y,z)])$

2.87 $\mathscr{S}$ has the form $(\exists x)(\exists y)(\forall z)([A_1^2(x,y) \Rightarrow A_1^1(x)] \Rightarrow A_1^1(z))$. Let the domain D be $\{1,2\}$, let A_1^2 be $<$, and let $A_1^1(u)$ stand for $u = 2$. Then $\mathscr{S}$ is true, but $(\forall x)(\exists y)A_1^2(x,y)$ is false.

2.88 Let g be a one-one correspondence between D^* and D. Define: $(a_j)^{M^*} = g((a_j)^M)$; $(f_j^n)^{M^*}(b_1,\ldots,b_n) = g^{-1}[(f_j^n)^M(g(b_1),\ldots,g(b_n))]$; $\models_{M^*} A_j^n[b_1,\ldots,b_n]$ if and only if $\models_M A_j^n[g(b_1),\ldots,g(b_n)]$.

2.95 *Hint*: Extend K by adding axioms $\mathscr{B}_n$, where $\mathscr{B}_n$ asserts that there are at least n elements. The new theory has no finite models.

2.96 (a) *Hint*: Consider the wfs $\mathscr{B}_n$, where $\mathscr{B}_n$ asserts that there are at least n elements. Use elimination of quantifiers, treating the $\mathscr{B}_n$s as if they were atomic wfs.

2.101 Let W be any set. For each b in W, let a_b be an individual constant. Let the theory K have as its proper axioms: $a_b \neq a_c$ for all b, c in W such that $b \neq c$, plus the axioms for a total order. K is consistent, since any finite subset of its axioms has a model. (Any such finite subset contains only a finite number of individual constants. One can define a total order on any finite set B by using the one–one correspondence between B and a set $\{1, 2, 3, \ldots, n\}$ and carrying over to B the total order $<$ on $\{1, 2, 3, \ldots, n\}$.) Since K is consistent, K has a model M by the generalized completeness theorem. The domain D of M is totally ordered by the relation $<^{\mathrm{M}}$; hence, the subset D_w of D consisting of the objects $(a_b)^{\mathrm{M}}$ is totally ordered by $<^{\mathrm{M}}$. This total ordering of D_w can then be carried over to a total ordering of W: $b <_w c$ if and only if $a_b <^{\mathrm{M}} a_c$.

2.103 Assume M_1 is finite and $\mathrm{M}_1 \equiv \mathrm{M}_2$. Let the domain D_1 of M_1 have n elements. Then, since the assertion that a model has exactly n elements can be written as a sentence, the domain D_2, of M_2 must also have n elements. Let $D_1 = \{b_1, \ldots, b_n\}$ and $D_2 = \{c_1, \ldots, c_n\}$.

Assume M_1 and M_2 are not isomorphic. Let φ be any one of the $n!$ one–one correspondences between D_1 and D_2. Since φ is not an isomorphism, either: (1) there is an individual constant a and an element b_j of D_1 such that either (i) $b_j = a^{\mathrm{M}_1} \wedge \varphi(b_j) \neq a^{\mathrm{M}_2}$ or (ii) $b_j \neq a^{\mathrm{M}_1} \wedge \varphi(b_j) = a^{\mathrm{M}_2}$; or (2) there is a function letter f_k^m and $b_\ell, b_{j_1}, \ldots, b_{j_m}$ in D_1 such that

$$b_\ell = (f_k^m)^{\mathrm{M}_1}(b_{j_1}, \ldots, b_{j_m}) \text{ and } \varphi(b_\ell) \neq (f_k^m)^{\mathrm{M}_2}(\varphi(b_{j_1}), \ldots, \varphi(b_{j_m}))$$

or (3) there is a predicate letter A_k^m and $b_{j_1}, \ldots, b_{j_m}$ in D_1 such that either

(i) $\models_{\mathrm{M}_1} A_k^m[b_{j_1}, \ldots, b_{j_m}]$ and $\models_{\mathrm{M}_2} \neg A_k^m[\varphi(b_{j_1}), \ldots, \varphi(b_{j_m})]$ or

(ii) $\models_{\mathrm{M}_1} \neg A_k^m[b_{j_1}, \ldots, b_{j_m}]$ and $\models_{\mathrm{M}_2} A_k^m[\varphi(b_{j_1}), \ldots, \varphi(b_{j_m})]$. Construct a wf $\mathscr{B}_\varphi$ as follows:

$$\mathscr{B}_\varphi \text{ is } \begin{cases} x_j = a & \text{if (1) (i) holds} \\ x_j \neq a & \text{if (1) (ii) holds} \\ x_\ell = f_k^m(x_{j_1}, \ldots, x_{j_m}) & \text{if (2) holds} \\ A_k^m(x_{j_1}, \ldots, x_{j_m}) & \text{if (3) (i) holds} \\ \neg A_k^m(x_{j_1}, \ldots, x_{j_m}) & \text{if (3) (ii) holds} \end{cases}$$

Let $\varphi_1, \ldots, \varphi_{n!}$ be the one–one correspondences between D_1 and D_2. Let $\mathscr{A}$ be the wf

$$(\exists x_1) \ldots (\exists x_n)\left(\bigwedge_{1 \leq i < j \leq n} x_i \neq x_j \wedge \mathscr{B}_{\varphi_1} \wedge \mathscr{B}_{\varphi_2} \wedge \cdots \wedge \mathscr{B}_{\varphi_{n!}} \right)$$

Then $\mathscr{A}$ is true for M_1 but not for M_2.

2.104 (a) There are $\aleph_\alpha$ sentences in the language of K. Hence, there are $2^{\aleph_\alpha}$ sets of sentences. If $M_1 \equiv M_2$ does not hold, then the set of sentences true for M_1 is different from the set of sentences true for M_2.

2.105 Let K^* be the theory with $\aleph_\gamma$ new symbols b_τ and, as axioms, all sentences true for M and all $b_\tau \neq b_\rho$ for $\tau \neq \rho$. Prove K^* consistent and apply Corollary 2.34.

2.108 (a) Let M be the field of rational numbers and let $X = \{-1\}$.

2.110 Consider the wf $(\exists x_2)\ x_2 < x_1$.

2.111 (a) (ii) Introduce a new individual constant b and form a new theory by adding to the complete diagram of M_1 all the sentences $b \neq t$ for all closed terms t of the language of K.

2.112 If $\emptyset \notin \mathscr{F}$, $\mathscr{F} \neq \mathscr{P}(A)$. Conversely, if $\emptyset \in \mathscr{F}$, then, by clause (3) of the definition of *filter*, $\mathscr{F} = \mathscr{P}(A)$.

2.113 If $\mathscr{F} = \mathscr{F}_B$, then $\bigcap_{C \in \mathscr{F}} C = B \in \mathscr{F}$. Conversely, if $B = \bigcap_{C \in \mathscr{F}} C \in \mathscr{F}$, then $\mathscr{F} = \mathscr{F}_B$.

2.114 Use Exercise 2.113.

2.115 (a) $A \in \mathscr{F}$, since $A = A - \emptyset$.

(b) If $B = A - W_1 \in \mathscr{F}$ and $C = A - W_2 \in \mathscr{F}$, where W_1 and W_2 are finite, then $B \cap C = A - (W_1 \cup W_2) \in \mathscr{F}$, since $W_1 \cup W_2$ is finite.

(c) If $B = A - W \in \mathscr{F}$, where W is finite, and if $B \subseteq C$, then $C = A - (W - C) \in \mathscr{F}$, since $W - C$ is finite.

(d) Let $B \subseteq C$. So, $B = A - W$, where W is finite. Let $b \in B$. Then $W \cup \{b\}$ is finite. Hence, $C = A - (W \cup \{b\}) \in \mathscr{F}$. But, $B \not\subseteq C$, since $b \notin C$. Therefore, $\mathscr{F} \neq \mathscr{F}_B$.

2.118 Let $\mathscr{F}' = \{D | D \subseteq A \wedge (\exists C)(C \in \mathscr{F} \wedge B \cap C \subseteq D)\}$.

2.119 Assume that, for every $B \subseteq A$, either $B \in \mathscr{F}$ or $A - B \in \mathscr{F}$. Let $\mathscr{G}$ be a filter such that $\mathscr{F} \subset \mathscr{G}$. Let $B \in \mathscr{G} - \mathscr{F}$. Then $A - B \in \mathscr{F}$. Hence, $A - B \in \mathscr{G}$. So, $\emptyset = B \cap (A - B) \in \mathscr{G}$ and $\mathscr{G}$ is improper. The converse follows from Exercise 2.118.

2.120 Assume $\mathscr{F}$ is an ultrafilter and $B \notin \mathscr{F}, C \notin \mathscr{F}$. By Exercise 2.119, $A - B \in \mathscr{F}$ and $A - C \in \mathscr{F}$. Hence, $A - (B \cup C) = (A - B) \cap (A - C) \in \mathscr{F}$. Since $\mathscr{F}$ is proper, $B \cup C \notin \mathscr{F}$. Conversely, assume $B \notin \mathscr{F} \wedge C \notin \mathscr{F} \Rightarrow B \cup C \notin \mathscr{F}$. Since $B \cup (A - B) = A \in \mathscr{F}$, this implies that, if $B \notin \mathscr{F}$, then $A - B \in \mathscr{F}$. Use Exercise 2.119.

2.121 (a) Assume $\mathscr{F}_C$ is a principal ultrafilter. Let $a \in C$ and assume $C \neq \{a\}$. Then $\{a\} \notin \mathscr{F}_C$ and $C - \{a\} \notin \mathscr{F}_C$. By Exercise 2.120, $C = \{a\} \cup (C - \{a\}) \notin \mathscr{F}_C$, which yields a contradiction.

(b) Assume a non-principal ultrafilter $\mathscr{F}$ contains a finite set, and let B be a finite set in $\mathscr{F}$ of least cardinality. Since $\mathscr{F}$ is non-principal, the cardinality of B is greater than 1. Let $b \in B$. Then $B - \{b\} \neq \emptyset$. Both $\{b\}$ and $B - \{b\}$ are finite sets of lower cardinality than B. Hence, $\{b\} \notin \mathscr{F}$ and $B - \{b\} \notin \mathscr{F}$. By Exercise 2.120, $B = \{b\} \cup (B - \{b\}) \notin \mathscr{F}$, which contradicts the definition of B.

2.124 Let J be the set of all finite subsets of Γ. For each Δ in J, choose a model M_Δ of Δ. For Δ in J, let $\Delta^* = \{\Delta' \mid \Delta' \in J \wedge \Delta \subseteq \Delta'\}$. The collection $\mathscr{G}$ of all Δ^*s has the finite-intersection property. By Exercise 2.117, there is a proper filter $\mathscr{F} \supseteq \mathscr{G}$. By the ultrafilter theorem, there is an ultrafilter $\mathscr{F}' \supseteq \mathscr{F} \supseteq \mathscr{G}$. Consider $\prod_{\Delta \in J} M_\Delta / \mathscr{F}'$. Let $\mathscr{B} \in \Gamma$. Then $\{\mathscr{B}\}^* \in \mathscr{G} \subseteq \mathscr{F}'$. Therefore, $\{\mathscr{B}\}^* \subseteq \{\Delta \mid \Delta \in \mathscr{G} \wedge \models_{M_\Delta} \mathscr{B}\} \in \mathscr{F}'$. By Loś's theorem, $\mathscr{B}$ is true in $\prod_{\Delta \in J} M_\Delta / \mathscr{F}'$.

2.125 (a) Assume $\mathscr{W}$ is closed under elementary equivalence and ultraproducts. Let Δ be the set of all sentences of $\mathscr{L}$ that are true in every interpretation in $\mathscr{W}$. Let M be any model of Δ. We must show that M is in $\mathscr{W}$. Let Γ be the set of all sentences true for M. Let J be the set of finite subsets of Γ. For $\Gamma' = \{\mathscr{B}_1, \ldots, \mathscr{B}_n\} \in J$, choose an interpretation $N_{\Gamma'}$ in $\mathscr{W}$ such that $\mathscr{B}_1 \wedge \ldots \wedge \mathscr{B}_n$ is true in $N_{\Gamma'}$. (If there were no such interpretation, $\neg(\mathscr{B}_1 \wedge \ldots \wedge \mathscr{B}_n)$, though false in M, would be in Δ.) As in Exercise 2.124, there is an ultrafilter $\mathscr{F}'$ such that $N^* = \prod_{\Gamma' \in J} N_{\Gamma'} / \mathscr{F}'$ is a model of Γ. Now, $N^* \in \mathscr{W}$. Moreover, $M \equiv N^*$. Hence, $M \in \mathscr{W}$.

(b) Use (a) and Exercise 2.59.

(c) Let $\mathscr{W}$ be the class of all fields of characteristic 0. Let $\mathscr{F}$ be a nonprincipal ultrafilter on the set P of primes, and consider $M = \prod_{p \in P} Z_p / \mathscr{F}$. Apply (b).

2.126 $R^\# \subseteq R^*$. Hence, the cardinality of R^* is $\geq 2^{\aleph_0}$. On the other hand, R^ω is equinumerous with 2^ω and, therefore, has cardinality $2^{\aleph_0}$. But the cardinality of R^* is at most that of R^ω.

2.127 Assume x and y are infinitesimals. Let ε be any positive real. Then $|x| < \varepsilon/2$ and $|y| < \varepsilon/2$. So, $|x+y| \leqslant |x| + |y| < \varepsilon/2 + \varepsilon/2 = \varepsilon$; $|xy| = |x||y| < 1 \cdot \varepsilon = \varepsilon$; $|x-y| \leqslant |x| + |-y| < \varepsilon/2 + \varepsilon/2 = \varepsilon$.

2.128 Assume $|x| < r_1$ and $|y| < \varepsilon$ for all positive real ε. Let ε be a positive real. Then ε/r_1 is a positive real. Hence $|y| < \varepsilon/r_1$, and so, $|xy| = |x||y| < r_1(\varepsilon/r_1) = \varepsilon$.

2.130 Assume $x - r_1$ and $x - r_2$ are infinitesimals, with r_1 and r_2 real. Then $(x - r_1) - (x - r_2) = r_2 - r_1$ is infinitesimal and real. Hence, $r_2 - r_1 = 0$.

2.131 (a) $x - \mathrm{st}(x)$ and $y - \mathrm{st}(y)$ are infinitesimals. Hence, their sum $(x+y) - (\mathrm{st}(x) + \mathrm{st}(y))$ is an infinitesimal. Since $\mathrm{st}(x) + \mathrm{st}(y)$ is real, $\mathrm{st}(x) + \mathrm{st}(y) = \mathrm{st}(x+y)$ by Exercise 2.130.

2.132 (a) By Proposition 2.45, $s^*(n) \approx c_1$ and $u^*(n) \approx c_2$ for all $n \in \omega^* - \omega$. Hence, $s^*(n) + u^*(n) \approx c_1 + c_2$ for all $n \in \omega^* - \omega$. But $s^*(n) + u^*(n) = (s+u)^*(n)$. Apply Proposition 2.45.

2.133 Assume f continuous at c. Take any positive real ε. Then there is a positive real δ such that $(\forall x)(x \in B \wedge |x - c| < \delta \Rightarrow |f(x) - f(c)| < \varepsilon)$ holds in $\mathscr{R}$. Therefore, $(\forall x)(x \in B^* \wedge |x - c| < \delta \Rightarrow |f^*(x) - f(c)| < \varepsilon)$ holds in $\mathscr{R}^*$. So, if $x \in B^*$ and $x \approx c$, then $|x - c| < \delta$ and, therefore, $|f^*(x) - f(c)| < \varepsilon$. Since ε was arbitrary, $f^*(x) \approx f(c)$. Conversely, assume $x \in B^* \wedge x \approx c \Rightarrow f^*(x) \approx f(c)$. Take any positive real ε. Let δ_0 be a positive

infinitesimal. Then $(\forall x)(x \in B^* \wedge |x - c| < \delta_0 \Rightarrow |f^*(x) - f(c)| < \varepsilon)$ holds for $\mathscr{R}^*$. Hence, $(\exists\delta)(\delta > 0 \wedge (\forall x)(x \in B^* \wedge |x - c| < \delta \Rightarrow |f'(x) - f(c)| < \varepsilon))$ holds for $\mathscr{R}^*$, and so, $(\exists\delta)(\delta > 0 \wedge (\forall x)(x \in B \wedge |x - c| < \delta \Rightarrow |f(x) - f(c)| < \varepsilon))$ holds in $\mathscr{R}$.

2.134 (a) Since $x \in B^* \wedge x \approx c \Rightarrow (f^*(x) \approx f(c) \wedge g^*(x) \approx g(c))$ by Proposition 2.46, we can conclude $x \in B^* \wedge x \approx c \Rightarrow (f + g)^*(x) \approx (f + g)(c)$, and so, by Proposition 2.46, $f + g$ is continuous at c.

2.139 (a)

(i)	$\neg[(\forall x)(A_1^1(x) \vee A_2^1(x)) \Rightarrow ((\forall x)A_1^1(x)) \vee (\forall x)A_2^1(x)]$	
(ii)	$(\forall x)(A_1^1(x) \vee A_2^1(x))$	(i)
(iii)	$\neg[((\forall x)A_1^1(x)) \vee (\forall x)A_2^1(x)]$	(i)
(iv)	$\neg(\forall x)A_1^1(x)$	(iii)
(v)	$\neg(\forall x)A_2^1(x)$	(iii)
(vi)	$(\exists x)\neg A_1^1(x)$	(iv)
(vii)	$(\exists x)\neg A_2^1(x)$	(v)
(viii)	$\neg A_1^1(b)$	(vi)
(ix)	$\neg A_2^1(c)$	(vii)
(x)	$A_1^1(b) \vee A_2^1(b)$	(ii)
(xi)	$A_1^1(b)$	(x)
(xii)	$\times$	
	$A_2^1(b)$	(x)
	$A_1^1(c) \vee A_2^1(c)$	(ii)
(xiii)	$A_1^1(c)$	(xii)
	$A_2^1(c)$	(xii)
	$\times$	

No further rules are applicable and there is an unclosed branch. Let the model M have domain $\{b, c\}$, let $(A_1^1)^M$ hold only for c, and let $(A_2^1)^M$ hold for only b. Then, $(\forall x)(A_1^1(x) \vee A_2^1(x))$ is true for M, but $(\forall x)A_1^1(x)$ and $(\forall x)A_2^1(x)$ are both false for M. Hence, $(\forall x)(A_1^1(x) \vee A_2^1(x)) \Rightarrow ((\forall x)A_1^1(x)) \vee (\forall x)A_2^1(x)$ is not logically valid.

CHAPTER 3

3.4 Consider the interpretation that has as its domain the set of polynomials with integral coefficients such that the leading coefficient is non-negative. The usual operations of addition and multiplication are the interpretations of $+$ and $\cdot$. Verify that (S1)–(S8) hold but that Proposition 3.11 is false (substituting the polynomial x for x and 2 for y).

3.5 (a) Form a new theory S' by adding to S a new individual constant b and the axioms $b \neq 0, b \neq \bar{1}, b \neq \bar{2}, \ldots, b \neq \bar{n}, \ldots$. Show that S′ is consistent, and apply Proposition 2.26 and Corollary 2.34(c).

(b) By a *cortège* let us mean any denumerable sequence of 0s and 1s. There are $2^{\aleph_0}$ cortèges. An element c of a denumerable model M of S determines a cortège $(s_0, s_1, s_2, \ldots)$ as follows: $s_i = 0$ if $\models_M p_i|c$, and $s_i = 1$ if $\models_M \neg(p_i|c)$. Consider now any cortège s. Add a new constant b to S, together with the axioms $\mathscr{B}_i(b)$, where $\mathscr{B}_i(b)$ is $p_i|b$ if $s_i = 0$ and $\mathscr{B}_i(b)$ is $\neg(p_i|b)$ if $s_i = 1$. This theory is consistent and, therefore, has a denumerable model M_s, in which the interpretation of b determines the cortège s. Thus, each of the $2^{\aleph_0}$ cortèges is determined by an element of some denumerable model. Every denumerable model determines denumerably many cortèges. Therefore, if a maximal collection of mutually non-isomorphic denumerable models had cardinality $\mathfrak{m} < 2^{\aleph_0}$, then the total number of cortèges represented in all denumerable models would be $\leqslant \mathfrak{m} \times \aleph_0 < 2^{\aleph_0}$. (We use the fact that the elements of a denumerable model determine the same cortèges as the elements of an isomorphic model.)

3.6 Let $(D, 0, ')$ be one model of Peano's postulates, with $0 \in D$ and $'$ the successor operation, and let $(D\#, 0\#, ^*)$ be another such model. For each x in D, by an *x-mapping* we mean a function f from $S_x = \{u | u \in D \wedge u \leqslant x\}$ into $D\#$ such that $f(0) = 0\#$ and $f(u') = (f(u))^*$ for all $u < x$. Show by induction that, for every x in D, there is a unique x-mapping (which will be denoted f_x). It is easy to see that, if $x_1 < x_2$, then the restriction of f_{x_2} to S_{x_1} must be f_{x_1}. Define $F(x) = f_x(x)$ for all x in D. Then F is a function from D into $D\#$ such that $F(0) = 0\#$ and $F(x') = (F(x))^*$ for all x in D. It is easy to prove that F is one–one. (If not, a contradiction results when we consider the least x in D for which there is some y in D such that $x \neq y$ and $F(x) = F(y)$.) To see that F is an isomorphism, it only remains to show that the range of F is $D\#$. If not, let z be the least element of $D\#$ not in the range of F. Clearly, $z \neq 0\#$. Hence, $z = w^*$ for some w. Then w is in the range of F, and so $w = F(u)$ for some u in D. Therefore, $F(u') = (F(u))^* = w^* = z$, contradicting the fact that z is not in the range of F.

The reason why this proof does not work for models of first-order number theory S is that the proof uses mathematical induction and the least-number principle several times, and these uses involve properties that cannot be formulated within the language of S. Since the validity of mathematical induction and the least-number principle in models of S is guaranteed to hold, by virtue of axiom (S9), only for wfs of S, the categoricity proof is not applicable. For example, in a non-standard model for S, the property of being the interpretation of one of the standard integers $\bar{0}, \bar{1}, \bar{2}, \bar{3}, \ldots$ is not expressible by a wf of S. If it were, then, by axiom (S9), one could prove that $\{\bar{0}, \bar{1}, \bar{2}, \bar{3}, \ldots\}$ constitutes the whole model.

3.7 Use a reduction procedure similar to that given for the theory K_2 on pages 116–17. For any number k, define $k \cdot t$ by induction: $0 \cdot t$ is 0 and $(k+1) \cdot t$ is $(k \cdot t) + t$; thus, $k \cdot t$ is the sum of t taken k times. Also, for any given k, let $t \equiv s(\mathrm{mod}\, k)$ stand for $(\exists x)(t = s + k \cdot x \vee s = t + k \cdot x)$. In the reduction procedure, consider all such wfs $t \equiv s(\mathrm{mod}\, k)$, as well as the wfs $t < s$, as atomic wfs, although they actually are not. Given any wfs of S_+, we may assume by Proposition 2.30 that it is in prenex normal form. Describe a method that, given a wf $(\exists y)\mathscr{C}$, where $\mathscr{C}$ contains no quantifiers (remembering the convention that $t \equiv s(\mathrm{mod}\, k)$ and $t < s$ are considered atomic), finds an equivalent wf without quantifiers (again remembering our convention). For help on details, see Hilbert and Bernays (1934, I, pp. 359–366).

3.8 (b) Use part (a) and Proposition 3.6(a)(i).

(c) Use part (b) and Lemma 1.12.

3.13 Assume $f(x_1, \ldots, x_n) = x_{n+1}$ is expressible in S by $\mathscr{B}(x_1, \ldots, x_{n+1})$. Let $\mathscr{C}(x_1, \ldots, x_{n+1})$ be $\mathscr{B}(x_1, \ldots, x_{n+1}) \wedge (\forall z)(z < x_{n+1} \Rightarrow \neg \mathscr{B}(x_1, \ldots, x_{n+1}))$. Show that $\mathscr{C}$ represents $f(x_1, \ldots, x_n)$ in S. [Use Proposition 3.8(b).] Assume, conversely, that $f(x_1, \ldots, x_n)$ is representable in S by $\mathscr{A}(x_1, \ldots, x_{n+1})$. Show that the same wf expresses $f(x_1, \ldots, x_n) = x_{n+1}$ in S.

3.16 (a) $(\exists y)_{u<y<v}\mathrm{R}(x_1, \ldots, x_n, y)$ is equivalent to $(\exists z)_{z<v \dot{-} (u+1)}\mathrm{R}(x_1, \ldots, x_n, z + u + 1)$, and similarly for the other cases.

3.18 If the relation $\mathrm{R}(x_1, \ldots, x_n, y)$: $f(x_1, \ldots, x_n) = y$ is recursive, then C_{R} is recursive and, therefore, so is $f(x_1, \ldots, x_n) = \mu y(C_{\mathrm{R}}(x_1, \ldots, x_n, y) = 0)$. Conversely, if $f(x_1, \ldots, x_n)$ is recursive, $C_{\mathrm{R}}(x_1, \ldots, x_n, y) = \mathrm{sg}|f(x_1, \ldots, x_n) - y|$ is recursive.

3.19

$$[\sqrt{n}] = \delta(\mu y_{y \leqslant n+1}(y^2 > n))$$

$$\Pi(n) = \sum_{y \leqslant n} \overline{\mathrm{sg}}(C_{\mathrm{Pr}}(y))$$

3.20 $[ne] = \left[n\left(1 + 1 + \frac{1}{2!} + \frac{1}{3!} + \cdots + \frac{1}{n!}\right)\right]$, since $n\left(\frac{1}{(n+1)!} + \frac{1}{(n+2)!} + \cdots\right) < \frac{1}{n!}$.

Let $1 + 1 + \frac{1}{2!} + \cdots + \frac{1}{n!} = \frac{g(n)}{n!}$. Then $g(0) = 1$ and $g(n+1) = (n+1)g(n) + 1$. Hence, g is primitive recursive. Therefore, so is $[ne] = \left[\frac{ng(n)}{n!}\right] = \mathrm{qt}(n!, ng(n))$.

3.21 $\mathrm{RP}(y, z)$ stands for $(\forall x)_{x \leqslant y+z}(x|y \wedge x|z \Rightarrow x = 1)$.

$$\varphi(n) = \sum_{y \leqslant n} \overline{\mathrm{sg}}(C_{\mathrm{RP}}(y, n))$$

3.22 $Z(0) = 0, Z(y+1) = U_2^2(y, Z(y))$.

3.23 Let $v = (p_0 p_1 \ldots p_k) + 1$. Some prime q is a divisor of v. Hence, $q \leqslant v$. But q is different from $p_0, p_1, \ldots, p_k$. If $q = p_j$, then $p_j|v$ and $p_j|p_0 p_1 \ldots p_k$ would imply that $p_j|1$ and, therefore, $p_j = 1$. Thus, $p_{k+1} \leqslant q \leqslant (p_0 p_1 \ldots p_k) + 1$.

3.26 If Goldbach's conjecture is true, h is the constant function 2. If Goldbach's conjecture is false, h is the constant function 1. In either case, h is primitive recursive.

3.28 List the recursive functions step by step in the following way. In the first step, start with the finite list consisting of $Z(x), N(x)$, and $U_1^1(x)$. At the $(n+1)$th step, make one application of substitution, recursion and the μ-operator to all appropriate sequences of functions already in the list after the nth step, and then add the $n+1$ functions $U_j^{n+1}(x_1, \ldots, x_{n+1})$ to the list. Every recursive function eventually appears in the list.

3.29 Assume $f_x(y)$ is primitive recursive (or recursive). Then so is $f_x(x)+1$. Hence, $f_x(x)+1$ is equal to $f_k(x)$ for some k. Therefore, $f_k(x) = f_x(x)+1$ for all x and, in particular, $f_k(k) = f_k(k)+1$.

3.30 (a) Let d be the least positive integer in the set Y of integers of the form $au+bv$, where u and v are arbitrary integers – say, $d = au_0 + bv_0$. Then $d|a$ and $d|b$. (To see this for a, let $a = qd + r$, where $0 \leqslant r < d$. Then $r = a - qd = a - q(au_0 + bv_0) = (1 - qu_0)a + (-qv_0)b \in Y$. Since d is the least positive integer in Y and $r < d, r$ must be 0. Hence $d|a$.) If a and b are relatively prime, then $d = 1$. Hence, $1 = au_0 + bv_0$. Therefore, $au_0 \equiv 1 \pmod{b}$.

3.32 (a) $1944 = 2^3 3^5$. Hence, 1944 is the Gödel number of the expression ( ).

(b) $49 = 1 + 8(2^1 3^1)$. Hence, 49 is the Gödel number of the function letter f_1^1.

3.34 (a) $g(f_1^1) = 49$ and $g(a_1) = 15$. So, $g(f_1^1(a_1)) = 2^{49}3^3 5^{15} 7^5$.

3.37 Take as a normal model for RR, but not for S, the set of polynomials with integral coefficients such that the leading coefficient is non-negative. Note that $(\forall x)(\exists y)(x = y+y \vee x = y+y+1)$ is false in this model but is provable in S.

3.38 Let ∞ be an object that is not a natural number. Let $\infty' = \infty, \infty + x = x + \infty = \infty$ for all natural numbers $x, \infty \cdot 0 = 0 \cdot \infty = 0$, and $\infty \cdot x = x \cdot \infty = \infty$ for all $x \neq 0$.

3.41 Assume S is consistent. By Proposition 3.37(a), $\mathscr{G}$ is not provable in S. Hence, by Lemma 2.12, the theory S_g is consistent. Now, $\neg\mathscr{G}$ is equivalent to $(\exists x_2)\mathscr{P}f(x_2, \ulcorner \mathscr{G} \urcorner)$. Since there is no proof of $\mathscr{G}$ in S, $\mathscr{P}f(k, q)$ is false for all natural numbers k, where $q = \ulcorner \mathscr{G} \urcorner$. Hence, $\vdash_S \neg \mathrm{Pf}(\bar{k}, \bar{q})$ for all natural numbers k. Therefore, $\vdash_{S_g} \neg\mathscr{P}f(\bar{k}, \bar{q})$. But, $\models_{S_g} (\exists x_2)\mathscr{P}f(x_2, \bar{q})$. Thus S_g is ω-inconsistent.

3.45 (G. Kreisel, *Mathematical Reviews*, 1955, Vol. 16, p. 103) Let $\mathscr{B}(x_1)$ be a wf of S that is the arithmetization of the following: x_1 is the Gödel number of a closed wf $\mathscr{B}$ such that the theory $S + \{\mathscr{B}\}$ is ω-inconsistent. (The latter says that there is a wf $\mathscr{E}(x)$ such that, for every n, $\mathscr{E}(\bar{n})$ is provable in $S + \{\mathscr{B}\}$, and such that $(\exists x)\neg\mathscr{E}(x)$ is provable in $S + \{\mathscr{B}\}$.) By the fixed-point theorem, let $\mathscr{C}$ be a closed wf such that $\vdash_S \mathscr{C} \Longleftrightarrow \mathscr{B}(\ulcorner \mathscr{C} \urcorner)$. Let $K = S + \{\mathscr{C}\}$. (1) $\mathscr{C}$ is false in the standard model. (Assume $\mathscr{C}$ true. Then K

is a true theory. But, $\mathscr{C} \Longleftrightarrow \mathscr{B}(\ulcorner \mathscr{C} \urcorner)$ is true, since it is provable in S. So, $\mathscr{B}(\ulcorner \mathscr{C} \urcorner)$ is true. Hence, K is ω- inconsistent and, therefore, K is not true, which yields a contradiction.) (2) K is ω-consistent. (Assume K ω-inconsistent. Then $\mathscr{B}(\ulcorner \mathscr{C} \urcorner)$ is true and, therefore, $\mathscr{C}$ is true, contradicting (1).)

3.46 (a) Assume the 'function' form of Church's thesis and let A be an effectively decidable set of natural numbers. Then the characteristic function C_A is effectively computable and, therefore, recursive. Hence, by definition, A is a recursive set.

(b) Assume the 'set' form of Church's thesis and let $f(x_1, \ldots, x_n)$ be any effectively computable function. Then the relation $f(x_1, \ldots, x_n) = y$ is effectively decidable. Using the functions σ^k, σ_i^k of pages 183–4 let A be the set of all z such that $f(\sigma_1^{n+1}(z), \ldots, \sigma_n^{n+1}(z)) = \sigma_{n+1}^{n+1}(z)$. Then A is an effectively decidable set and, therefore, recursive. Hence, $f(x_1, \ldots, x_n) = \sigma_{n+1}^{n+1}(\mu z(C_A(z) = 0))$ is recursive.

3.48 Let K be the extension of S that has as proper axioms all wfs that are true in the standard model. If *Tr* were recursive, then, by Proposition 3.38, K would have an undecidable sentence, which is impossible.

3.49 Use Corollary 3.39.

3.50 Let $f(x_1, \ldots, x_n)$ be a recursive function. So, $f(x_1, \ldots, x_n) = y$ is a recursive relation, expressible in K by a wf $\mathscr{A}(x_1, \ldots, x_n, y)$. Then f is representable by $\mathscr{A}(x_1, \ldots, x_n, y) \wedge (\forall z)(z < y \Longrightarrow \neg \mathscr{A}(x_1, \ldots, x_n, z))$, where $z < y$ stands for $z \leqslant y \wedge z \neq y$.

3.53 (a) $\vdash 0 = 1 \Longrightarrow \mathscr{G}$. Hence, $\vdash \mathscr{Bew}(\ulcorner 0 = \bar{1} \urcorner) \Longrightarrow \mathscr{Bew}(\ulcorner \mathscr{G} \urcorner)$ and, therefore, $\vdash \neg \mathscr{Bew}(\ulcorner \mathscr{G} \urcorner) \Longrightarrow \neg \mathscr{Bew}(\ulcorner 0 = \bar{1} \urcorner)$. Thus, $\vdash \mathscr{G} \Longrightarrow \neg \mathscr{Bew}(\ulcorner 0 = \bar{1} \urcorner)$.

(b) $\vdash \mathscr{Bew}(\ulcorner \mathscr{G} \urcorner) \Longrightarrow \mathscr{Bew}(\ulcorner \mathscr{Bew}(\ulcorner \mathscr{G} \urcorner) \urcorner)$. Also, $\vdash \neg \mathscr{G} \Longleftrightarrow \mathscr{Bew}(\ulcorner \neg \mathscr{G} \urcorner)$, and so, $\vdash \mathscr{Bew}(\ulcorner \neg \mathscr{G} \urcorner) \Longleftrightarrow \mathscr{Bew}(\ulcorner \mathscr{Bew}(\ulcorner \mathscr{G} \urcorner) \urcorner)$. Hence $\vdash \mathscr{Bew}(\ulcorner \mathscr{G} \urcorner) \Longrightarrow \mathscr{Bew}(\ulcorner \neg \mathscr{G} \urcorner)$. By a tautology, $\vdash \mathscr{G} \Longrightarrow (\neg \mathscr{G} \Longrightarrow (\mathscr{G} \wedge \neg \mathscr{G}))$; hence, $\vdash \mathscr{Bew}(\ulcorner \mathscr{G} \urcorner) \Longrightarrow \mathscr{Bew}(\ulcorner \neg \mathscr{G} \Longrightarrow (\mathscr{G} \wedge \neg \mathscr{G}) \urcorner)$. Therefore, $\vdash \mathscr{Bew}(\ulcorner \mathscr{G} \urcorner) \Longrightarrow (\mathscr{Bew}(\ulcorner \neg \mathscr{G} \urcorner) \Longrightarrow \mathscr{Bew}(\ulcorner (\mathscr{G} \wedge \neg \mathscr{G}) \urcorner))$. It follows that $\vdash \mathscr{Bew}(\ulcorner \mathscr{G} \urcorner) \Longrightarrow \mathscr{Bew}(\ulcorner (\mathscr{G} \wedge \neg \mathscr{G}) \urcorner)$. But, $\vdash \mathscr{G} \wedge \neg \mathscr{G} \Longrightarrow 0 = \bar{1}$; so, $\vdash \mathscr{Bew}(\ulcorner (\mathscr{G} \wedge \neg \mathscr{G}) \urcorner) \Longrightarrow \mathscr{Bew}(\ulcorner 0 = \bar{1} \urcorner)$. Thus, $\vdash \mathscr{Bew}(\ulcorner \mathscr{G} \urcorner) \Longrightarrow \mathscr{Bew}(\ulcorner 0 = \bar{1} \urcorner)$, and $\vdash \neg \mathscr{Bew}(\ulcorner 0 = \bar{1} \urcorner) \Longrightarrow \neg \mathscr{Bew}(\ulcorner \mathscr{G} \urcorner)$. Hence, $\vdash \neg \mathscr{Bew}(\ulcorner 0 = \bar{1} \urcorner) \Longrightarrow \mathscr{G}$.

3.56 If a theory K is recursively decidable, the set of Gödel numbers of theorems of K is recursive. Taking the theorems of K as axioms, we obtain a recursive axiomatization.

3.58 Assume there is a recursive set C such that $T_K \subseteq C$ and $\mathrm{Ref}_K \subseteq \overline{C}$. Let C be expressible in K by $\mathscr{A}(x)$. Let $\mathscr{F}$, with Gödel number k, be a fixed point for $\neg \mathscr{A}(x)$. Then, $\vdash_K \mathscr{F} \Longleftrightarrow \neg \mathscr{A}(\bar{k})$. Since $\mathscr{A}(x)$ expresses C in K, $\vdash_K \mathscr{A}(\bar{k})$ or $\vdash_K \neg \mathscr{A}(\bar{k})$.

(a) If $\vdash_K \mathscr{A}(k)$, then $\vdash_K \neg\mathscr{F}$. Therefore, $k \in \mathrm{Ref}_K \subseteq \overline{C}$. Hence, $\vdash_K \neg\mathscr{A}(\bar{k})$, contradicting the consistency of K.

(b) If $\vdash_K \neg\mathscr{A}(\bar{k})$, then $\vdash_K \mathscr{F}$. So, $k \in T_K \subseteq C$ and therefore, $\vdash_K \mathscr{A}(\bar{k})$, contradicting the consistency of K.

3.60 Let K_2 be the theory whose axioms are those wfs of K_1 that are provable in K^*. The theorems of K_2 are the axioms of K_2. Hence, $x \in T_{K_2}$ if and only if $\mathrm{Fml}_{K_1}(x) \wedge x \in T_{K^*}$. So, if K^* were recursively decidable – that is, if T_{K^*} were recursive – T_{K_2} would be recursive. Since K_2 is a consistent extension of K_1, this would contradict the essential recursive undecidability of K_1.

3.61 (a) Compare the proof of Proposition 2.28.

(b) By part (a), K^* is consistent. Hence, by Exercise 3.60, K^* is essentially recursively undecidable. So, by (a), K is recursively undecidable.

3.62 (b) Take $(\forall x)(A_j^1(x) \Longleftrightarrow x = x)$ as a possible definition of A_j^1.

3.63 Use Exercises 3.61(b) and 3.62.

3.64 Use Corollary 3.46, Exercise 3.63, and Proposition 3.47.

CHAPTER 4

4.12 (s) Assume $u \in x \times y$. Then $u = \langle v, w\rangle = \{\{v\}, \{v, w\}\}$ for some v in x and w in y. Then $v \in x \cup y$ and $w \in x \cup y$. So, $\{v\} \in \mathscr{P}(x \cup y)$ and $\{v, w\} \in \mathscr{P}(x \cup y)$. Hence, $\{\{v\}, \{v, w\}\} \in \mathscr{P}(\mathscr{P}(x \cup y))$.

4.15 (a) $\mathscr{D}(x) \subseteq \bigcup(\bigcup x)$ and $\mathscr{R}(x) \subseteq \bigcup(\bigcup x)$. Apply Corollary 4.6(b).

(b) Use Exercise 4.12(s), Exercise 4.13(b), axiom W, and Corollary 4.6(b).

(c) If $\mathrm{Rel}(Y)$, then $Y \subseteq \mathscr{D}(Y) \times \mathscr{R}(Y)$. Use part (b) and Corollary 4.6(b).

4.18 Let $X = \{\langle y_1, y_2\rangle | y_1 = y_2 \wedge y_1 \in Y\}$; that is, X is the class of all ordered pairs $\langle u, u\rangle$ with $u \in Y$. Clearly, $\mathrm{Fnc}(X)$ and, for any set x, $(\exists v)(\langle v, u\rangle \in X \wedge v \in x) \Longleftrightarrow u \in Y \cap x$. So, by axiom R, M(Y ∩ x)

4.19 Assume $\mathrm{Fnc}(Y)$. Then $\mathrm{Fnc}(x\restriction Y)$ and $\mathscr{D}(x\restriction Y) \subseteq x$. By axiom R, M ($Y$"$x$).

4.22 (a) Let $\emptyset$ be the class $\{u | u \neq u\}$. Assume $\mathrm{M}(X)$. Then $\emptyset \subseteq X$. So, $\emptyset = \emptyset \cap X$. By axiom S, M ($\emptyset$).

4.23 Assume $\mathrm{M}(V)$. Let $Y = \{x | x \notin x\}$. It was proved above that $\neg\mathrm{M}(Y)$. But $Y \subseteq V$. Hence, by Corollary 4.6(b), $\neg\mathrm{M}(V)$.

4.30 (c) Let u be the least $\in$-element of $X - Z$.

4.33 (a) By Proposition 4.11(a), $\mathrm{Trans}(\omega)$. By Proposition 4.11(b) and Proposition 4.8(j), $\omega \in On$. If $\omega \in K_1$ then $\omega \in \omega$, contradicting Proposition 4.8(a). Hence, $\omega \notin K_1$.

4.39 Let $X_1 = X \times \{\emptyset\}$ and $Y_1 = Y \times \{1\}$.

4.40 For any $u \subseteq y$, let the characteristic function C_u be the function with domain y such that $C_u{}'w = \emptyset$ if $w \in u$ and $C_u{}'w = 1$ if $w \in y - u$. Let F be the function with domain $\mathscr{P}(y)$ such that $F'u = C_u$ for $u \in \mathscr{P}(y)$. Then $\mathscr{P}(x) \underset{F}{\cong} 2^y$

4.41 (a) For any set u, $\mathscr{D}(u)$ is a set by Exercise 4.15(a).

(b) If $u \in x^y$, then $u \subseteq y \times x$. So, $x^y \subseteq \mathscr{P}(y \times x)$.

4.42 (a) $\emptyset$ is the only function with domain $\emptyset$.

(c) If $\mathscr{D}(u) \neq \emptyset$, then $\mathscr{R}(u) \neq \emptyset$.

4.43 Define a function F with domain X such that, for any x_0 in X, $F(x_0)$ is the function g in $X^{\{u\}}$ such that $g'u = x_0$. Then $X \underset{F}{\cong} X^{\{u\}}$.

4.44 Assume $X \underset{F}{\cong} Y$ and $Z \underset{G}{\cong} W$. If $\neg \mathrm{M}(W)$, then $\underset{F}{\neg} \mathrm{M}(Z)$ and $X^Z = Y^W = \emptyset$ by Exercise 4.41(a). Hence, we may assume $\mathrm{M}(W)$ and $\mathrm{M}(Z)$. Define a function Φ on X^Z as follows: if $f \in X^Z$, let $\Phi'f = F \circ f \circ G^{-1}$. Then $X^Z \underset{\Phi}{\cong} Y^W$.

4.45 If X or Y is not a set, then $Z^{X \cup Y}$ and $Z^X \times Z^Y$ are both $\emptyset$. We may assume then that X and Y are sets. Define a function Φ with domain $Z^{X \cup Y}$ as follows: if $f \in Z^{X \cup Y}$, let $\Phi'f = \langle X \restriction f, Y \restriction f \rangle$. Then $Z^{X \cup Y} \underset{\Phi}{\cong} Z^X \times Z^Y$.

4.46 Define a function F with domain $(x^y)^z$ as follows: for any f in $(x^y)^z$, let $F'f$ be the function in $x^{y \times z}$ such that $(F'f)'\langle u, v \rangle = (f'v)'u$ for all $\langle u, v \rangle \in y \times z$. Then $(x^y)^z \underset{F}{\cong} x^{y \times z}$.

4.47 If $\neg \mathrm{M}(Z)$, $(X \times Y)^Z = \emptyset = \emptyset \times \emptyset = X^Z \times Y^Z$. Assume then that $\mathrm{M}(Z)$. Define a function F: $X^Z \times Y^Z \to (X \times Y)^Z$ as follows: for any $f \in X^Z, g \in Y^Z, (F'\langle f, g \rangle)'z = \langle f'z, g'z \rangle$ for all z in Z. Then $X^Z \times Y^Z \underset{F}{\cong} (X \times Y)^Z$.

4.48 This is a direct consequence of Proposition 4.19.

4.54 (b) Use Bernstein's theorem (Proposition 4.23(d)).

(c) Use Proposition 4.23(c,d).

4.55 Define a function F from V into 2_c as follows: $F'u = \{u, \emptyset\}$ if $u \neq \emptyset$; $F'\emptyset = \{1, 2\}$. Since, F is one-one, $V \preccurlyeq 2_c$. Hence, by Exercises 4.23 and 4.50, $\neg \mathrm{M}(2_c)$.

4.56 (h) Use Exercise 4.45.

(i) $2^x \preccurlyeq 2^x +_c x \preccurlyeq 2^x +_c 2^x = 2^x \times 2 \cong 2^x \times 2^1 \cong 2^{x+_c 1} \cong 2^x$.

Hence, by Bernstein's Theorem, $2^x +_c x \cong 2^x$.

4.59 Under the assumption of the axiom of infinity, ω is a set such that $(\exists u)(u \in \omega) \wedge (\forall y)(y \in \omega \Rightarrow (\exists z)(z \in \omega \wedge y \subset z))$. Conversely, assume (*) and let b be a set such that (i) $(\exists u)(u \in b)$ and (ii) $(\forall y)(y \in b \Rightarrow (\exists z)(z \in b \wedge y \subset z))$. Let $d = \{u | (\exists z)(z \in b \wedge u \subseteq z)\}$. Since $d \subseteq \mathscr{P}(\bigcup(b))$, d is a set. Define a relation $R = \{\langle n, v \rangle | n \in \omega \wedge v = \{u | u \in d \wedge u \cong n\}\}$. Thus, $\langle n, v \rangle \in R$ is and only if $n \in \omega$ and v consists of all elements of d that are equinumerous with n. R is a one–one function with domain ω and range a subset of $\mathscr{P}(d)$. Hence, by the replacement axiom applied to R^{-1}, ω is a set and, therefore, axiom I holds.

4.62 (a) Induction on α in $(\forall x)(x \cong \alpha \wedge \alpha \in \omega \Rightarrow \mathrm{Fin}(\mathscr{P}(x)))$.

(b) Induction on α in $(\forall x)(x \cong \alpha \wedge \alpha \in \omega \wedge (\forall y)(y \in x \Rightarrow \mathrm{Fin}(y)) \Rightarrow \mathrm{Fin}(\bigcup x))$.
(c) Use Proposition 4.27(a).
(d) $x \subseteq \mathscr{P}(\bigcup x)$ and $y \in x \Rightarrow y \subseteq \bigcup x$.
(e) Induction on α in $(\forall x)(x \cong \alpha \wedge \alpha \in \omega \Rightarrow (x \preccurlyeq y \vee y \preccurlyeq x))$
(g) Induction on α in $(\forall x)(x \cong \alpha \wedge \alpha \in \omega \wedge \mathrm{Inf}(Y) \Rightarrow x \preccurlyeq Y)$
(h) Use Proposition 4.26(c).
(j) $x^y \subseteq \mathscr{P}(y \times x)$

4.63 Let Z be a set such that every non-empty set of subsets of Z has a minimal element. Assume Inf(Z). Let Y be the set of all infinite subsets of Z. Then Y is a non-empty set of subsets of Z without a minimal element. Conversely, prove by induction that, for all α in ω, any non-empty subset of $\mathscr{P}(\alpha)$ has a minimal element. The result then carries over to non-empty subsets of $\mathscr{P}(z)$, where z is any finite set.

4.64 (a) Induction on α in $(\forall x)(x \cong \alpha \wedge \alpha \in \omega \wedge \mathrm{Den}(y) \Rightarrow \mathrm{Den}(x \cup y))$.
(b) Induction on α in $(\forall x)(x \cong \alpha \wedge x \neq \emptyset \wedge \mathrm{Den}(y) \Rightarrow \mathrm{Den}(x \times y))$
(c) Assume $z \subseteq x$ and Den(z). Let $z \cong \omega$. Define a function g on x as follows: $g`u = u$ if $u \in x - z$; $g`u \overset{f}{=} (\breve{f})`((f`u)')$ if $u \in z$. Assume x is Dedekind-infinite. Assume $z \subset x$ and $x \cong z$. Let $v \in x - z$. Define a function h on ω such that $h`\emptyset = v$ and $h`(\alpha') = f`(h`\alpha)$ if $\alpha \in \omega$. Then h is one–one. So, Den($h``\omega$) and $h``\omega \subseteq x$.
(f) Assume $y \notin x$. (i) Assume $x \cup \{y\} \cong x$. Define by induction a function g on ω such that $g`\emptyset = y$ and $g`(n+1) = f`(g`n)$. g is a one–one function from ω into x. Hence, x contains a denumerable subset and, by part (c), x is Dedekind-infinite. (ii) Assume x is Dedekind-infinite. Then, by part (c), there is a denumerable subset z of x. Assume $z \cong \omega$. Let $c_0 = (f^{-1})`\emptyset$. Define a function $F$ as follows: $F`u = u$ for $u \overset{f}{\in} x - z$; $F`c_0 = y$; $F`u = (f^{-1})`((f`u - 1)$ for $u \in z - \{c_0\}$. Then $x \underset{F}{\cong} x \cup \{y\}$. If z is $\{c_0, c_1, c_2, \ldots\}$, F takes c_{i+1} into c_i and moves c_0 into y.
(g) Assume $\omega \preccurlyeq x$. By part (c), x is Dedekind-infinite. Choose $y \notin x$. By part (f), $x \cong x \cup \{y\}$. Hence, $x +_c 1 = (x \times \{\emptyset\}) \cup \{\langle \emptyset, 1 \rangle\}$ $\cong x \cup \{y\} \cong x$.

4.65 Assume M is a model of NBG with denumerable domain D. Let z be the element of D satisfying the wf $x = 2^\omega$. Hence, z satisfies the wf $\neg(x \cong \omega)$. This means that there is no object in D that satisfies the condition of being a one–one correspondence between z and ω. Since D is denumerable, there is a one–one correspondence between the set of 'elements' of z (that is, the set of objects v in D such that $\models_M v \in z$) and the set of natural numbers. However, no such one–one correspondence exists within M.

4.68 NBG is finitely axiomatizable and has only the binary predicate letter A_2^2. The argument on p. 269–70 shows that NBG is recursively undecidable. Hence, by Proposition 3.49, the predicate calculus with A_2^2 as its only non-logical constant is recursively undecidable.

4.69 (a) Assume $x \preccurlyeq \omega_\alpha$. If $2 \preccurlyeq x$, then, by Propositions 4.37(b) and 4.40, $\omega_\alpha \preccurlyeq x \cup \omega_\alpha \preccurlyeq x \times \omega_\alpha \preccurlyeq \omega_\alpha \times \omega_\alpha \cong \omega_\alpha$. If x contains one element, use Exercise 4.64(c,f).

(b) Use Corollary 4.41.

4.70 (a) $\mathscr{P}(\omega_\alpha) \times \mathscr{P}(\omega_\alpha) \cong 2^{\omega_\alpha} \times 2^{\omega_\alpha} \cong 2^{\omega_\alpha +_c \omega_\alpha} \cong 2^{\omega_\alpha} \cong \mathscr{P}(\omega_\alpha)$

(b) $(\mathscr{P}(\omega_\alpha))^x \cong (2^{\omega_\alpha})^x \cong 2^{\omega_\alpha \times x} \cong 2^{\omega_\alpha} \cong \mathscr{P}(\omega_\alpha)$

4.71 (a) If y were non-empty and finite, $y \cong y +_c y$ would contradict Exercise 4.62(b).

(b) By part (c), let $y = u \cup v, u \cap v = \emptyset, u \cong y, v \cong y$. Let $y \cong v$. Define a function g on $\mathscr{P}(y)$ as follows: for $x \subseteq y$, let $g`x = u \cup f``x)$. Then $g`x \subseteq y$ and $y \cong u \preccurlyeq g`x \preccurlyeq y$. Hence, $g`x \cong y$. So, g is a one–one function from $\mathscr{P}(y)$ into $A = \{z | z \subseteq y \wedge z \cong y\}$. Thus, $\mathscr{P}(y) \preccurlyeq A$. Since $A \subseteq \mathscr{P}(y)$, $A \preccurlyeq \mathscr{P}(y)$.

(c) Use part (d): $\{z | z \subseteq y \wedge z \cong y\} \subseteq \{z | z \subseteq y \wedge \mathrm{Inf}(z)\}$.

(f) By part (c), let $y = u \cup v, u \cap v = \emptyset, u \cong y, v \cong y$. Let $u \underset{h}{\cong} v$. Define f on y as follows: $f`x = h`x$ if $x \in u$ and $f`x = (h^{-1})`x$ if $x \in v$.

4.72 (a) Use Proposition 4.37(b).

(b) (i) Perm $(y) \subseteq y^y \preccurlyeq (2^y)^y \cong 2^{y \times y} \cong 2^y \cong \mathscr{P}(y)$.

(ii) By part (a), we may use Exercise 4.7 (c). Let $y = u \cup v, u \cap v = \emptyset$, $u \cong y, v \cong y$. Let $u \underset{H}{\cong} v$ and $y \underset{G}{\cong} u$. Define a function $F: \mathscr{P}(y) \to$ Perm (y) in the following way: assume $z \in \mathscr{P}(y)$. Let $\psi_z : y \to y$ be defined as follows: $\psi_z`x = H`x$ if $x \in G``z$; $\psi_z`x = (H^{-1})`x$ if $(H^{-1})`x \in G``z$; $\psi_z`x = x$ otherwise. Then $\psi_z \in$ Perm$(y)$. Let $F`z = \psi_z$. F is one–one. Hence, $\mathscr{P}(y) \preccurlyeq Perm(y)$.

4.73 (a) Use WO and Proposition 4.19.

(b) The proof of Zorn $\Rightarrow$ WO in Proposition 4.42 uses only this special case of Zorn's Lemma.

(c) To prove the Hausdorff maximal principal (HMP) from Zorn, consider some $\subset$-chain C_0 in x. Let y be the set of all $\subset$-chains C in x such that $C_0 \subseteq C$ and apply part (b) to y. Conversely, assume HMP. To prove part (b), assume that the union of each non-empty $\subset$-chain in a given non-empty set x is also in x. By HMP applied to the $\subset$-chain $\emptyset$, there is some maximal $\subset$-chain C in x. Then $\bigcup(C)$ is an $\subset$-maximal element of x.

(d) Assume the Teichmüller–Tukey lemma (TT). To prove part (b), assume that the union of each non-empty $\subset$-chain in a given non-empty set x is also in x. Let y be the set of all $\subset$-chains in x. y is easily seen to be a set of finite character. Therefore, y contains a $\subset$-maximal element C. Then $\bigcup(C)$ is a $\subset$-maximal element of x. Conversely, let x be any set of finite character. In order to prove TT by means of part (b), we must show that, if C is a $\subset$-chain in x, then $\bigcup(C) \in x$. By the finite character of x, it suffices to show that every finite subset z of $\bigcup(C)$ is in x. Now, since z is finite, z is a

subset of the union of a finite subset W of C. Since C is a $\subset$-chain, W has a $\subset$-greatest element $w \in x$, and z is a subset of w. Since x is of finite character, $z \in x$.

(e) Assume Rel(x). Let $u = \{z | (\exists v)(v \in \mathscr{D}(x) \wedge z = \{v\}\{x\}\}$; that is, $z \in u$ if z is the set of all ordered pairs $\langle v, w\rangle$ in x, for some fixed v. Apply the multiplicative axiom to u. The resulting choice set $y \subseteq x$ is a function with domain $\mathscr{D}(x)$. Conversely, the given property easily yields the multiplicative axiom. If x is a set of disjoint non-empty sets, let r be the set of all ordered pairs $\langle u, v\rangle$ such that $u \in x$ and $v \in u$. By part (e), there is a function $f \subseteq r$ such that $\mathscr{D}(f) = \mathscr{D}(r) = x$. The range $\mathscr{R}(f)$ is the required choice set for x.

(f) By trichotomy, either $x \prec y$ or $y \prec x$. If $x \prec y$, there is a function with domain y and range x. (Assume $x \overset{f}{\cong} y_1 \subseteq y$. Take $c \in x$. Define $g`u = c$ if $u \in y - y_1$, and $g`u = (f^{-1})`u$ if $u \in y_1$.) Similarly, if $y \prec x$, there is a function with domain $x$ and range $y$. Conversely, to prove WO, apply the assumption (f) to $x$ and $\mathscr{H}`(\mathscr{P}(x))$. Note that, if $(\exists f)(f : u \to v \wedge \mathscr{R}(f) = v)$, then $\mathscr{P}(v) \prec \mathscr{P}(u)$. Therefore, if there were a function f from x onto $\mathscr{H}`(\mathscr{P}(x))$, we would have $\mathscr{H}`(\mathscr{P}(x)) \prec \mathscr{P}(\mathscr{H}`(\mathscr{P}(x))) \prec \mathscr{P}(x)$ contradicting the definition of $\mathscr{H}`(\mathscr{P}(x))$. Hence, there is a function from $\mathscr{H}`(\mathscr{P}(x))$ onto $x$. Since $\mathscr{H}`(\mathscr{P}(x))$ is an ordinal, one can define a one-one function from x into $\mathscr{H}`(\mathscr{P}(x))$. Thus $x \prec \mathscr{H}`(\mathscr{P}(x))$ and, therefore, x can be well-ordered.

4.76 If $<$ is a partial ordering of x, use Zorn's lemma to obtain a maximal partial ordering $<^*$ of x with $< \subseteq <^*$. But a maximal partial ordering must be a total ordering. (If u, v were distinct elements of x unrelated by $<^*$, we could add to $<^*$ all pairs $\langle u_1, v_1\rangle$ such that $u_1 \leqslant^* u$ and $v \leqslant^* v_1$. The new relation would be a partial ordering properly containing $<^*$.)

4.79 (b) Since $x \times y \cong x +_c y$, $x \times y = a \cup b$ with $a \cap b = \phi$, $a \cong x$, $b \cong y$. Let r be a well-ordering of y. (i) Assume there exists u in x such that $\langle u, v\rangle \in a$ for all v in y. Then $y \preccurlyeq a$. Since $a \cong x$, $y \preccurlyeq x$, contradicting $\neg(y \preccurlyeq x)$. Hence, (ii) for any u in x, there exists v in y such that $\langle u, v\rangle \in b$. Define f: $x \to b$ such that $f`u = \langle u, v\rangle$, where v is the r-least element of y such that $\langle u, v\rangle \in b$. Since f is one–one, $x \preccurlyeq b \cong y$.

(c) Clearly Inf(z) and Inf($x +_c z$). Then

$$x +_c z \cong (x +_c z)^2 \cong x^2 +_c 2 \times (x \times z) +_c z^2 \cong x +_c 2 \times (x \times z) +_c z$$

Therefore, $x \times z \preccurlyeq 2 \times (x \times z) \preccurlyeq x +_c 2 \times (x \times z) +_c z \cong x +_c z$. Conversely, $x +_c z \preccurlyeq x \times z$ by Proposition 4.37(b).

(d) If AC holds, $(\forall y)(\text{Inf}(y) \Rightarrow y \cong y \times y)$ follows from Proposition 4.40 and Exercise 4.73(a). Conversely, if we assume $y \cong y \times y$ for all infinite y, then, by parts (c) and (b), it follows that $x \preccurlyeq \mathscr{H}`x$ for

any infinite set x. Since $\mathscr{H}$'x is an ordinal, x can be well-ordered. Thus, WO holds.

4.81 (a) Let $\langle$ be a well-ordering of the range of r. Let f'$\emptyset$ be the $\langle$-least element of $\mathscr{R}(r)$, and let f'n be the $\langle$-least element of those v in $\mathscr{R}(r)$ such that $\langle f'n, v\rangle \in r$.

(b) Assume Den(x) $\wedge$ $(\forall u)(u \in x \Rightarrow u \neq \emptyset)$. Let $\omega \underset{g}{\cong} x$. Let r be the set of all pairs $\langle a, b\rangle$ such that a and b are finite sequences $\langle v_0, v_1, \ldots v_n\rangle$ and $\langle v_0, v_1, \ldots, v_{n+1}\rangle$ such that, for $0 \leqslant i \leqslant n+1, v_i \in g'i$. Since $\mathscr{R}(r) \subseteq \mathscr{D}(r)$, PDC produces a function h: $\omega \to \mathscr{D}(r)$ such that $\langle h'n, h'(n')\rangle \in r$ for all n in ω. Define the choice function f by taking, for each u in x, $f'u$ to be the $(g'u)$th component of the sequence $h'(g'u)$.

(c) Assume PDC and Inf(x). Let r consist of all ordered pairs $\langle u, u\cup\{a\}\rangle$, where $u\cup\{a\} \subseteq x$, Fin$(u\cup\{a\})$, and $a \notin u$. By PDC, there is a function f: $\omega \to \mathscr{D}(r)$ such that$\langle f'n, f'(n')\rangle \in r$ for all n in ω. Define g: $\omega \to x$ by setting $g'n$ equal to the unique element of $f'(n')$–$f'n$. Then g is one–one, and so, $\omega \preccurlyeq x$.

(d) In the proof of Proposition 4.44(b), instead of using the choice function h, apply PDC to obtain the function f. As the relation r, use the set of all pairs $\langle u, v\rangle$ such that $u \in c, v \in c, v \in u\cap X$.

4.82 (a) Use transfinite induction.

(d) Use induction on β.

(e)–(f) Use transfinite induction and part (a).

(h) Assume $u \subseteq H$. Let v be the set of ranks $\rho'x$ of elements x in u. Let $\beta = \cup v$. Then $u \subseteq \Psi'\beta$. Hence $u \in \mathscr{P}(\Psi'\beta) = \Psi'(\beta) \subseteq H$..

4.83 Assume $X \neq \emptyset \wedge \neg(\exists y)(y \in X \wedge y\cap X = \emptyset)$. Choose $u \in X$. Define a function g: $g'\emptyset = u\cap X$, $g'(n') = \cup(g'n)\cap X$. Let $x = \cup(\mathscr{R}(g))$. Then $x \neq \emptyset$ and $(\forall y)(y \in x \Rightarrow y\cap x \neq \emptyset)$.

4.88 Hint: Assume that the other axioms of NBG are consistent and that the Axiom of Infinity is provable from them. Show that H_ω is a model for the other axioms but not for the Axiom of Infinity.

4.89 Use $H_{\omega +_o \omega}$

4.95 (a) Let $C = \{x|\ \neg(\exists y)(x \in y \wedge y \in x)\}$.

CHAPTER 5

5.1 $q_0|\mathbf{B}q_0$
$q_0\mathbf{B}\mathbf{R}q_1$
$q_1\|q_0$
$q_1\mathbf{B}\mathbf{R}q_2$

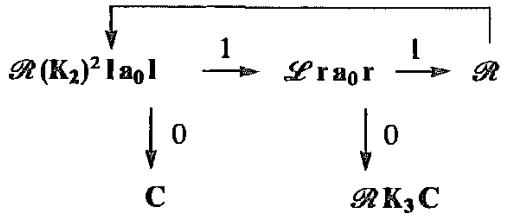
5.2 (a) U_2^3 (b) $\delta(x)$

5.7 Let a Turing machine $\mathscr{T}$ compute the function f. Replace all occurrences of q_0 in the quadruples of $\mathscr{T}$ by a new internal state q_r. Then add the

quadruples $q_0 a_i a_i q_r$ for all symbols a_i of the alphabet of $\mathscr{F}$. The Turing machine defined by the enlarged set of quadruples also computes the function f.

5.10 (a) $N(x) = x + 1$ (c) $2x$

5.12 (a)



5.14 (a) The empty function (b) $N(x) = x + 1$ (c) $Z(x)$

5.16 If $f(a_1) = b_1, \ldots, f(a_n) = b_n$, then

$$f(x) = \mu y[(x = a_1 \wedge y = b_1) \vee \cdots \vee (x = a_n \wedge y = b_n)]$$

5.20 Let $g(z,x) = U(\mu y T_1(z,x,y))$ and use Corollary 5.11. Let v_0 be a number such that $g(x,x) + 1 = g(v_0,x)$. Then, if $g(v_0,v_0)$ is defined, $g(v_0,v_0) + 1 = g(v_0,v_0)$, which is impossible.

5.21 $g(x_1,\ldots,x_n) = h_1(x_1,\ldots,x_n) \cdot \overline{\text{sg}}(C_{R_1}(x_1,\ldots,x_n)) + \ldots + h_k(x_1,\ldots,x_n) \cdot \overline{\text{sg}}(C_{R_k}(x_1,\ldots,x_n))$

5.22 (a) Assume that $h(x)$ is a recursive function such that $h(x) = \mu y T_1(x,x,y)$ for every x in the domain of $\mu y T_1(x,x,y)$. Then $(\exists y) T_1(x,x,y)$ if and only if $T_1(x,x,h(x))$. Since $T_1(x,x,h(x))$ is a recursive relation, this contradicts Corollary 5.13(a).

(b) Use Exercise 5.21.

(c) $Z(\mu y T_1(x,x,y))$ is recursively completable, but its domain is $\{x | (\exists y) T_1(x,x,y)\}$, which, by Corollary 5.13(a), is not recursive.

5.29 Let $\mathscr{T}$ be a Turing machine with a recursively unsolvable halting problem. Let a_k be a symbol not in the alphabet of $\mathscr{T}$. Let q_r be an internal state symbol that does not occur in the quadruples of $\mathscr{T}$. For each q_i of $\mathscr{T}$ and a_j of $\mathscr{T}$, if no quadruple of $\mathscr{T}$ begins with $q_i a_j$, then add the quadruple $q_i a_j a_k q_r$. Call the new Turing machine $\mathscr{T}^*$. Then, for any initial tape description α of $\mathscr{T}$, $\mathscr{T}^*$, begun on α, prints a_k if and only if $\mathscr{T}$ is applicable to α. Hence, if the printing problem for $\mathscr{T}^*$ and a_k were recursively solvable, then the halting problem for $\mathscr{T}$ would be recursively solvable.

5.31 Let $\mathscr{T}$ be a Turing machine with a recursively unsolvable halting problem. For any initial tape description α for $\mathscr{T}$, construct a Turing machine $\mathscr{T}_\alpha$ that does the following: for any initial tape description β, start $\mathscr{T}$ on α; if $\mathscr{T}$ stops, erase the result and then start $\mathscr{T}$ on β. It is easy to check that $\mathscr{T}$ is applicable to α if and only if $\mathscr{T}_\alpha$ has a recursively unsolvable halting problem. It is very tedious to show how to construct $\mathscr{T}_\alpha$ and to

prove that the Gödel number of $\mathscr{T}_\alpha$ is a recursive function of the Gödel number of α.

5.33 Let v_0 be the index of a partial recursive function $G(x)$ with non-empty domain. If the given decision problem were recursively solvable, so would be the decision problem of Example 1 on page 332

5.34 By Corollary 5.16, there is a recursive function $g(u)$ such that $\varphi^1_{g(u)}(x) = x \cdot \mu y T_1(u,u,y)$. Then $\varphi^1_{g(u)}$ has an empty domain if and only if $\neg(\exists y)T_1(u,u,y)$. But, $\neg(\exists y)T_1(u,u,y)$ is not recursive by Corollary 5.13(a).

5.39 (a) By Corollary 5.16, there is a recursive function $g(u)$ such that $\varphi^1_{g(u)}(x) = \mu y(x = u \wedge y = x)$. The domain of $\varphi^1_{g(u)}$ is $\{u\}$. Apply the fixed-point theorem to g.

(b) There is a recursive function $g(u)$ such that $\varphi^1_{g(u)}(x) = \mu y(x \neq u \wedge y = 0)$. Apply the fixed-point theorem to g.

5.42 (a) Let $A = \{x | f(x) \in B\}$. By Proposition 5.21(c), B is the domain of a partial recursive function g. Then A is the domain of the composition $g \circ f$. Since $g \circ f$ is partial recursive by substitution, A is r.e. by Proposition 5.21(c).

(b) Let B be a recursive set and let D be the inverse image of B under a recursive function f. Then $x \in D$ if and only if $C_B(f(x)) = 0$ and $C_B(f(x)) = 0$ is a recursive relation.

(c) Let B be an r.e. set and let A be the image $\{f(x) | x \in B\}$ under a partial recursive function f. If B is empty, so is A. If B is non-empty, then B is the range of a recursive function g. Then A is the range of the partial recursive function $f(g(x))$ and, by Proposition 5.21(b), A is r.e.

(d) Consider part (b). Given any natural number x, compute the value $f(x)$ and determine whether $f(x)$ is in B. This is an effective procedure for determining membership in the inverse image of B. Hence, by Church's thesis, B is recursive.

(e) Any non-empty r.e. set that is not recursive (such as that of Proposition 5.21(e)) is the range of a recursive function g and is, therefore, the image of the recursive set ω of all natural numbers under the function g.

5.43 The proof has two parts:

1. Let A be an infinite recursive set. Then A is the range of a recursive function f, by Proposition 5.21(d). Since A is infinite, $h(u) = \mu y(f(y) > u)$ is recursive. Let a_0 be the least element of A. Define $g(0) = a_0$, $g(n+1) = f(h(g(n))$. Then g is a strictly increasing function with range A.
2. Let A be the range of a strictly increasing recursive function g. Then $g(x) \geq x$ for all x (by the special case of Proposition 4.15). Hence, $x \in A$ if and only if $(\exists u)_{u \leq x} g(u) = x$. So, A is recursive by Proposition 3.18.

5.44 Assume A is an infinite r.e. set. Let A be the range of the recursive function $g(x)$. Define the function f by the following course-of-values recursion:

$$f(n) = g(\mu y((\forall z)_{z<n} g(y) \neq f(z))) = g(\mu y((\forall z)_{z<n} g(y) \neq (f\#(n))_z))$$

Then A is the range of h, h is one–one, and h is recursive by Propositions 3.18 and 3.20. Intuitively, $f(0) = g(0)$ and, for $n > 0, f(n) = g(y)$, where y is the least number for which $g(y)$ is different from $f(0), f(1), \ldots, f(n-1)$.

5.45 Let A be an infinite r.e. set, and let A be the range of the recursive function g. Since A is infinite, $F(u) = \mu y(g(y) > u)$ is a recursive function. Define $G(0) = g(0), G(n+1) = g(\mu y(g(y) > G(n))) = g(F(G(n)))$. G is a strictly increasing recursive function whose range is infinite and included in A. By Exercise 5.43, the range of G is an infinite recursive subset of A.

5.46 (a) By Corollary 5.16, there is a recursive function $g(u,v)$ such that $\varphi^1_{g(u,v)}(x) = \mu y(T_1(u,x,y) \vee T_1(v,x,y))$.

5.47 Assume (∇). Let $f(x_1, \ldots, x_n)$ be effectively computable. Then the set $B = \{u | f((u)_1, \ldots, (u)_n) = (u)_{n+1}\}$ is effectively enumerable and, therefore, by (∇), r.e. Hence, $u \in B \Longleftrightarrow (\exists y) R(u,y)$ for some recursive relation R. Then

$$f(x_1, \ldots, x_n) = ([\mu v(((v)_0)_1 = x_1 \wedge \cdots \wedge ((v)_0)_n = x_n \wedge R((v)_0, (v)_1))]_0)_{n+1}$$

So, f is recursive. Conversely, assume Church's thesis and let W be an effectively enumerable set. If W is empty, then W is r.e. If W is non-empty, let W be the range of the effectively computable function g. By Church's thesis, g is recursive. But, $x \in W \Longleftrightarrow (\exists u)(g(u) = x)$. Hence, W is r.e. by Proposition 5.21(a).

5.48 Assume A is r.e. Since $A \neq \emptyset$, A is the range of a recursive function $g(z)$. So, for each z, $U(\mu y T_1(g(z), x, y))$ is total and, therefore, recursive. Hence, $U(\mu y T_1(g(x), x, y)) + 1$ is recursive. Then there must be a number z_0 such that $U(\mu y T_1(g(x), x, y)) + 1 = U(\mu y T_1(g(z_0), x, y))$. A contradiction results when $x = z_0$.

5.49 (a) Let $\varphi(n) = n$ for all n.

5.50 Let $\varphi(z) = \sigma^2_1(\mu y[T_1(z, \sigma^2_1(y), \sigma^2_2(y)) \wedge \sigma^2_1(y) > 2z])$, and let B be the range of ϕ.

5.55 (b) Let A be r.e. Then $x \in A \Longleftrightarrow (\exists y) R(x,y)$, where R is recursive. Let $\mathscr{R}(x,y)$ express $R(x,y)$ in K. Then $k \in A \Longleftrightarrow \vdash_K (\exists y)\mathscr{R}(\bar{k}, y)$.

(c) Assume $k \in A \Longleftrightarrow \vdash_K \mathscr{A}(\bar{k})$ for all natural numbers k. Then $k \in A \Longleftrightarrow (\exists y) B_{\mathscr{A}}(k,y)$ and $B_{\mathscr{A}}$ is recursive (see the proof of Proposition 3.29 on page 199.

5.56 (a) Clearly T_K is infinite. Let $f(x)$ be a recursive function with range T_K. Let $\mathscr{B}_0, \mathscr{B}_1, \ldots$ be the theorems of K, where $\mathscr{B}_j$ is the wf of K with Gödel number $f(j)$. Let $g(x,y)$ be the recursive function such that, if x is the Gödel number of a wf $\mathscr{C}$, then $g(x,j)$ is the Gödel number of the conjunction $\mathscr{C} \wedge \mathscr{C} \wedge \cdots \wedge \mathscr{C}$ consisting of j conjuncts; and, otherwise, $g(x,j) = 0$. Then $g(f(j), j)$ is the Gödel

number of the j-fold conjunction $\mathscr{B}_j \wedge \mathscr{B}_j \wedge \cdots \wedge \mathscr{B}_j$. Let K′ be the theory whose axioms are all these j-fold conjunctions, for $j = 0, 1, 2, \ldots$ Then K′ and K have the same theorems. Moreover, the set of axioms of K′ is recursive. In fact, x is the Gödel number of an axiom of K′ if and only if $x \neq 0 \wedge (\exists y)_{y \leqslant x}(g(f(y), y) = x)$. From an intuitive standpoint using Church's thesis, we observe that, given any wf $\mathscr{A}$, one can decide whether $\mathscr{A}$ is a conjunction $\mathscr{C} \wedge \mathscr{C} \wedge \cdots \wedge \mathscr{C}$; if it is such a conjunction, one can determine the number j of conjuncts and check whether $\mathscr{C}$ is $\mathscr{B}_j$.

(b) Part (b) follows from part (a).

5.58 (a) Assume $\mathscr{B}(x_1)$ weakly expresses $(\bar{T}_K)^*$ in K. Then, for any n, $\vdash_K \mathscr{B}(\bar{n})$ if and only if $n \in (T_K)^*$. Let p be the Gödel number of $\mathscr{B}(x_1)$. Then $\vdash_K \mathscr{B}(\bar{p})$ if and only if $p \in (\bar{T}_K)^*$. Hence, $\vdash_K \mathscr{B}(\bar{p})$ if and only if the Gödel number of $\mathscr{B}(\bar{p})$ is in $\bar{T}_K$; that is, $\vdash_K \mathscr{B}(\bar{p})$ if and only if not-$\vdash_K \mathscr{B}(\bar{p})$.

(b) If K is recursively decidable, T_K is recursive. Hence, $\bar{T}_K$ is recursive and, by Exercise 5.57, $(\bar{T}_K)^*$ is recursive. So, $(\bar{T}_K)^*$ is weakly expressible in K, contradicting part (a).

(c) Use part (b); every recursive set is expressible, and, therefore, weakly expressible, in every consistent extension of K.

5.59 (a) (i) $\delta(x)$.

(ii) $x_1 \dot{-} x_2$

(iii) The function with empty domain.

(iv) The doubling function.

(b) (i)
$$f_1^2(x_1, 0) = x_1$$
$$f_1^2(0, x_2) = x_2$$
$$f_1^2((x_1)', (x_2)') = f_1^2(x_1, x_2)$$

(ii)
$$f_1^2(x_1, 0) = x_1$$
$$f_1^2(x_1, (x_2)') = (f_1^2(x_1, x_2))'$$
$$f_2^2(x_1, 0) = 0$$
$$f_2^2(x_1, (x_2)') = f_1^2(f_2^2(x_1, x_2), x_1)$$

(iii)
$$f_1^1(0) = \bar{1}$$
$$f_1^1((x_1)') = 0$$
$$f_2^1(0) = 0$$
$$f_2^1((x_1)') = f_1^1(f_2^1(x_1))$$

5.61 (a) Any word P is transformed into QP.

(b) Any word P in A is transformed into PQ.

(c) Any word P in A is transformed into Q.

(d) Any word P in A is transformed into $\bar{n}$, where n is the number of symbols in P.

5.62 (a) $\alpha\xi \to \cdot\ \Lambda \quad (\xi \text{ in A})$
$\alpha \to \cdot\ \Lambda$
$\Lambda \to \alpha$

(b) $\alpha\xi \to \xi\alpha \quad (\xi \text{ in A})$
$\xi\alpha \to \cdot\ \Lambda \quad (\xi \text{ in A})$
$\alpha \to \cdot\ \Lambda$
$\Lambda \to \alpha$

(c) $\xi \to \Lambda \quad (\xi \text{ in A})$
$\alpha\alpha \to \cdot\ \Lambda$
$\Lambda \to \cdot\ \alpha$

(d) $\xi\eta\beta \to \eta\beta\xi \quad (\xi, \eta \text{ in A})$
$\alpha\xi \to \xi\beta\xi\alpha \quad (\xi \text{ in A})$
$\beta \to \gamma$
$\gamma \to \Lambda$
$\alpha \to \cdot\ \Lambda$
$\Lambda \to \alpha$

5.63 $\alpha a_i \to Q_i\alpha \quad (i = 1, \ldots, k)$
$\alpha\xi \to \xi\alpha \quad (\xi \text{ in A} - \{a_1, \ldots, a_k\})$
$\alpha \to \cdot\ \Lambda$
$\Lambda \to \alpha$

5.64 (d) $|\mathrm{B}| \to \mathrm{B}$
$\mathrm{B} \to |$

(e) $|\mathrm{B}| \to |$

(f) Let α, β and δ be new symbols.

$$\begin{aligned}
\beta| &\to |\beta \\
\alpha| &\to |\beta\alpha \\
\alpha &\to \Lambda \\
||\delta &\to |\delta\alpha \\
|\delta &\to | \\
\delta|| &\to \delta| \\
\delta| &\to | \\
\delta &\to | \\
\beta &\to | \\
|\mathrm{B}| &\to \delta
\end{aligned}$$

Bibliography

Listed here are not only books and papers mentioned in the text but also some other material that will be helpful in a further study of mathematical logic. We shall use the following abbreviations.

AML for *Annals of Mathematical Logic*
AMS for American Mathematical Society
Arch. for *Archiv für mathematische Logik und Grundlagenforschung*
FM for *Fundamenta Mathematicae*
HML for *Handbook of Mathematical Logic*, Springer-Verlag
HPL for *Handbook of Philosophical Logic*, Reidel
JSL for *Journal of Symbolic Logic*
MTL for *Model-Theoretic Logics*, Springer-Verlag
NDJFL for *Notre Dame Journal of Formal Logic*
NH for North-Holland Publishing Company
ZML for *Zeitschrift für mathematische Logik und Grundlagen der Mathematik* (since 1993, *Mathematical Logic Quarterly*)

Ackermann, W. (1928) Zum Hilbertschen Aufbau der reellen Zahlen, *Math. Annalen*, 99, 118–133. (1937) Die Widerspruchsfreiheit der allgemeinen Mengenlehre, *ibid.*, 114, 305–315. (1940) Zur Widerspruchsfreiheit der Zahlentheorie, *ibid.*, 117, 162–194. (1954) *Solvable Cases of the Decision Problem*, NH.

Andrews, P. (1965) *Transfinite Type Theory with Type Variables*, NH. (1986) *An Introduction to Mathematical Logic and Type Theory. To Truth Through Proof*, Academic.

Barwise, J. (1985) Model-theoretic logics: Background and aims, *MTL*, 3–23.

Barwise, J. and S. Feferman (eds) (1985) *Model-Theoretic Logics*, Springer.

Baudisch, A., D. Seese, P. Tuschik and M. Weese (1985) Decidability and quantifier elimination, *MTL*, 235–268

Behmann, H. (1992) Beiträge zur Algebra der Logik, insbesondere zum Entscheidungsproblem, *Math. Annalen*, 86, 163–229.

Bell, J.L. (1977) *Boolean-Valued Models and Independence Proofs in Set Theory*, Oxford University Press.

Bernardi, C. (1975) The fixed-point theorem for diagonalizable algebras, *Studia Logica*, 34, 239–252. (1976) The uniqueness of the fixed-point in every diagonalizable algebra, *ibid.*, 35, 335–343.

Bernays, P. (1937–54) A system of axiomatic set theory, I, *JSL*, 2 (1937), 65–77; II, 6 (1941), 1–17; III, 7 (1942), 65–89; IV, 7 (1942), 133–145; V, 8 (1943), 89–104; VI, 13 (1948), 65–79; VII, 19 (1954), 81–96. (1957) Review of Myhill (1955), *JSL*, 22, 73–76. (1961) Zur Frage der Unendlichkeitsschemata in der axiomatischen Mengenlehre, *Essays on the Foundations of Mathematics*, Jerusalem, 3–49. (1976) On the problem of schemata of infinity in axiomatic set theory, *Sets and Classes*, NH.

Bernstein, A.R. (1973) Non-standard analysis, *Studies in Model Theory*, Math. Assoc. of America, 35 58.

Beth, E. (1951) A topological proof of the theorem of Löwenheim–Skolem–Gödel, *Indag. Math.*, 13, 436–444. (1953) Some consequences of the theorem of Löwenheim–Skolem–Gödel–Malcev, *ibid.*, 15, 66–71. (1959) *The Foundations of Mathematics*, NH.

Bezboruah, A. and J.C. Shepherdson (1976) Gödel's second incompleteness theorem for Q, *JSL*, 41, 503–512.

Bolc, L. and P. Borowik (1992) *Many-Valued Logics. Vol. I: Theoretical Foundations*, Springer.

Boolos, G. (1984) Trees and finite satisfiability: proof of a conjecture of Burgess, *NDJFL*, 25, 193–197. (1989) New proof of the Gödel incompleteness theorem, *Notices of the AMS*, 36, 388–390. (1993) *The Logic of Provability*, Cambridge University Press.

Boone, W. (1959) The word problem, *Ann. Math.*, 70, 207–265.

Bourbaki, N. (1947) *Algèbre*, Hermann, Paris, Book II, Chap. II.

Britton, J.L. (1963) The word problem, *Ann. Math.*, 77, 16–32.

Brouwer, L.E.J. (1976) *Collected Works, Vol. 1, Philosophy and Foundations of Mathematics*, NH.

Bruijn, N.G. de and P. Erdös (1951) A colour problem for infinite graphs and a problem in the theory of relations. *Indag. Math.*, 13, 369–373.

Brunner, N. (1990) The Fraenkel–Mostowski method revisited, *NDJFL*, 31, 64–75.

Carnap, R. (1934) *Die Logische Syntax der Sprache*, Springer (English translation, *The Logical Syntax of Language*, Routledge & Kegan Paul, 1937; text edition, Humanities, 1964.)

Chaitin, G. (1992) *Information-Theoretic Incompleteness*, World Scientific.

Chang, C.C. and H.J. Keisler (1973) *Model Theory*, second edition, NH (third edition, 1990).

Cherlin, G. (1976) *Model Theoretic Algebra, Selected Topics*, Springer.

Chuquai, R. (1972) Forcing for the impredicative theory of classes, *JSL*, 37, 1–18. (1981) *Axiomatic Set Theory. Impredicative Theories of Classes*, NH.

Church, A. (1936a) A note on the Entscheidungsproblem, *JSL*, 1, 40–41; correction, *ibid.*, 101–102 (reprinted in Davis, 1965). (1936b) An unsolvable problem of elementary number theory, *Am. J. Math.*, 58, 345–363 (reprinted in Davis, 1965). (1940) A formulation of the simple theory of types, *JSL*, 5, 56–68. (1941) *The Calculi of Lambda Conversion*, Princeton University Press (second printing, 1951). (1956) *Introduction to Mathematical Logic*, I, Princeton University Press.

Chwistek, L. (1924–25) The theory of constructive types, *Annales de la Soc. Polonaise de Math.*, 2, 9–48; 3, 92–141.

Cohen, P.J. (1963–64) The independence of the continuum hypothesis, *Proc. Natl. Acad. Sci. USA*, 50, 1143–1148; 51, 105–110. 1966. *Set Theory and the Continuum Hypothesis*, Benjamin.

Collins, G.E. (1955) The modeling of Zermelo set theories in New Foundations, Ph.D. thesis, Cornell.

Corcoran, J. (1980) Categoricity, *History and Philosophy of Logic*, 1, 187–207. (1987) Second-order logic, *Proceedings Inference OUIC 86* (eds D. Moates and R. Butrick), Ohio University Press, 7–31.

Craig, W. (1953) On axiomatizability within a system, *JSL*, 18, 30–32.

Curry, H.B. and R. Feys (1958) *Combinatory Logic, I*, NH.

Curry, H.B., J.R. Hindley and J. Seldin (1972) *Combinatory Logic, II*. NH.

Davis, M. (1958) *Computability and Unsolvability*, McGraw-Hill (Dover, 1983). (1965) (ed.). *The Undecidable: Basic Papers on Undecidable Propositions, Unsolvable Problems, and Computable Functions*, Raven. (1973) Hilbert's tenth problem is unsolvable, *Am. Math. Monthly*, 80, 233–269. (1977a) *Applied Nonstandard Analysis*, Wiley. (1977b) Unsolvable problems, *HML*, 567–594. (1982) Why Gödel didn't have Church's thesis, *Information and Control*, 54, 3–24.

Davis, M., H. Putnam and J. Robinson (1961) The decision problem for exponential Diophantine equations, *Annals of Math.*, 74, 425–436.

Dedekind, R. (1901) *Essays on the Theory of Numbers*. Open Court (Dover, 1963).

Dekker, J.C.E. (1953) Two notes on recursively enumerable sets, *Proc. AMS*, 4, 495–501. (1955) Productive sets, *Trans. AMS*, 78, 129–149.

Dekker, J.C.E. and J. Myhill (1960) Recursive equivalence types, *Univ. Calif. Publ. Math.*, 3, 67–213.

Denyer, N. (1991) Pure second-order logic, *NDJFL*, 33, 220–224.

Dreben, B. (1952) On the completeness of quantification theory, *Proc. Natl. Acad. Sci. USA*, 38, 1047–1052.

Dreben, B. and W.D. Goldfarb (1980) *Decision Problems, Solvable Classes of Quantificational Formulas*, Addison-Wesley.

Dummett, M. (1977) *Elements of Intuitionism*. Oxford University Press.

Easton, W.B. (1970) Powers of regular cardinals, *AML*, 1, 139–178.

Ehrenfeucht, A. (1957) On theories categorical in power, *FM*, 44, 241–248. (1958) Theories having at least continuum non-isomorphic models in each infinite power (abstract), *Notices AMS*, 5, 680.

Ehrenfeucht, A. and S. Feferman (1960) Representability of recursively enumerable sets in formal theories, *Arch.*, 5, 37–41.

Engeler, E. (1968) *Formal Languages: Automata and Structures*, Markham. (1973) *Introduction to the Theory of Computability*, Academic.

Erdös, P. and A. Tarski (1961) On some problems involving inaccessible cardinals, *Essays on the Foundations of Mathematics*, Magnes, Jerusalem, 50–82.

Ershov, Yu., I. Lavrov, A. Taimanov and M. Taitslin (1965) Elementary theories, *Russian Mathematical Surveys*, 20, 35–105.

Evans, T. (1951) The word problem for abstract algebras, *J. London Math. Soc.*, 26, 64–71.

Feferman, S. (1957) Degrees of unsolvability associated with classes of formalized theories, *JSL*, 22, 165–175. (1960) Arithmetization of metamathematics in a general setting, *FM*, 49, 35–92. (1962) Transfinite recursive progressions of axiomatic theories, *JSL*, 27, 259–316.

Felgner, U. (1971a) *Models of ZF-Set Theory*, Springer. (1971b) Comparisons of the axioms of local and universal choice, *FM*, 71, 43–62. (1976) Choice functions on sets and classes, *Sets and Classes*, NH, 217–255.

Fischer, P.C. (1965) On formalisms for Turing machines, *J. Assoc. Comp. Mach.*, 12, 570–580.

Fitting, M. (1983) *Proof Methods for Modal and Intuitionistic Logics*, Reidel.

Forster, T.E. (1983) *Quine's New Foundations (An Introduction)*, Cahiers du Centre de Logique 5, Université Catholique de Louvain. (1992) *Set Theory with a Universal Set: Exploring an Untyped Universe*, Oxford University Press.

Fraenkel, A.A. (1922a) Zu den Grundlagen der Cantor-Zermeloschen Mengenlehre, *Math. Annalen*, 86, 230–237. (1922b) Der Begriff 'definit' und die Unabhängigkeit des Auswahlaxioms, *Sitzungsberichte der Preussischen Akademie der Wissenschaften, Physikalisch-mathematische Klasse*, 253–257. (1928) *Einleitung in die Mengenlehre* (third edition), Springer.

Fraenkel, A., Y. Bar-Hillel, and A. Lévy. (1973) *Foundations of Set Theory*, NH (second revised edition).

Frayne, T., A. Morel and D. Scott. 1956. Reduced direct products, *FM*, 51, 195–228.

Frege, G. (1893, 1903) *Grundgesetze der Arithmetik, Begriffschriftlich Abgeleitet*, Vols 1–2, Jena (partial English translation in *The Basic Laws of Arithmetic: Exposition of the System*, University of California Press, 1964).

Gabriel, P. (1962) Des catégories abéliennes, *Bull. Soc. Math. France*, 90, 323–448.

Gandy, R. (1988) The confluence of ideas in 1936, *The Universal Turing Machine – a Half-century Survey* (ed. R. Herken), Oxford University Press, 55–111.

Garey, M.R. and D.S. Johnson. (1978) *Computers and Intractability: A Guide to the Theory of NP-Completeness*. Freeman.

Garland, S.J. (1974) Second-order cardinal characterizability, *Axiomatic Set Theory, Proc. of Symposia in Pure Mathematics*, Vol. XIII, Part II, American Mathematical Society, 127–146.

Gentzen, G. (1936) Die Widerspruchsfreiheit der reinen Zahlentheorie, *Math. Ann.*, 112, 493–565. (1938) Neue Fassung des Widerspruchsfreiheitsbeweises für die reine Zahlentheorie, *Forschungen zur Logik*, 4, 5–18. (1969) *Collected Papers* (ed. M.E. Szabo), NH.

Gillies, D.A. (1982) *Frege, Dedekind, and Peano on the Foundations of Arithmetic*, Van Gorcum. (1992) The Fregean revolution in logic, *Revolutions in Mathematics* (ed D.A. Gillies), Oxford University Press, 265–305.

Gödel, K. (1930) Die Vollständigkeit der Axiome des logischen Funktionenkalküls, *Monatsh. Math. Phys.*, 37, 349–360 (English translation in Van Heijenoort, 1967, 582–591). (1931) Ueber formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme, I, *ibid.*, 38, 173–198 (English translation in Davis, 1965). (1933) Zum intuitionistischen Aussagenkalkül; Zur intuitionistischen Arithmetik und Zahlentheorie, *Ergeb. Math. Koll.*, 4, 34–38 and 40 (translation in Davis, 1965). (1934) On undecidable propositions of formal mathematical systems, *Lecture Notes, Institute for Advanced Study*, Princeton University Press (reprinted in Davis, 1965, 39–73). (1938) The consistency of the axiom of choice and the generalized continuum hypothesis, *Proc. Natl. Acad. Sci. USA*, 24, 556–557. (1939) Consistency proof for the generalized continuum hypothesis, *ibid.*, 25, 220–226. (1940) *The Consistency of the Axiom of Choice and the Generalized Continuum Hypothesis with the Axioms of Set Theory*, Princeton University Press. (1947) What is Cantor's continuum problem? *Amer. Math. Monthly*, 54, 515–525. (1986, 1990, 1995) *Collected Works, Volume I, Publications 1929–1936. Volume II,*

Publications 1938–1974. Volume III, Unpublished Essays and Lectures, Oxford University Press.

Hailperin, T. (1944) A set of axioms for logic, *JSL*, 9, 1–19. (1953) Quantification theory and empty individual domains, *JSL*, 18, 197–200.

Hajek, P. (1993) *Metamathematics of First-Order Arithmetic*, Springer.

Hall, M., Jr (1949) The word problem for semigroups with two generators, *JSL*, 14, 115–118.

Halmos, P. (1960) *Naive Set Theory*, Van Nostrand (Springer, 1974). (1962) *Algebraic Logic*, Chelsea. (1963) *Lectures on Boolean Algebra*, Van Nostrand (Springer, 1977)

Halmos, P. and H. Vaughn (1950) The marriage problem, *Amer. J Math.*, 72, 214–215.

Halpern, J.D. (1964) The independence of the axiom of choice from the Boolean prime ideal theorem, *FM*, 55, 57–66.

Halpern, J.D. and A. Lévy (1971) The Boolean prime ideal theorem does not imply the axiom of choice, *Proc. Symp. in Pure Math.*, 13 AMS, 83–134.

Hartogs, F. (1915) Über das Problem der Wohlordnung, *Math. Annalen*, 76, 438–443.

Hasenjaeger, G. (1953) Eine Bemerkung zu Henkin's Beweis für die Vollständigkeit des Prädikatenkalküls der ersten Stufe, *JSL*, 18, 42–48.

Hasenjaeger, G. and H. Scholz (1961) *Grundzüge der mathematischen Logik*, Springer.

Hatcher, W. (1982) *The Logical Foundations of Mathematics*, Pergamon.

Heijenoort, J. van (1967) (ed.). *From Frege to Gödel (A Source Book in Mathematical Logic, 1879–1931)*. Harvard University Press.

Hellman, M. (1961) A short proof of an equivalent form of the Schröder Bernstein theorem, *Am. Math. Monthly*, 68, 770.

Henkin, L. (1949) The completeness of the first-order functional calculus, *JSL*, 14, 159–166. (1950) Completeness in the theory of types, *ibid.*, 15, 81–91. (1955) On a theorem of Vaught, *JSL*, 20, 92–93.

Henkin, L., J.D. Monk and A. Tarski. (1971, 1985) *Cylindric Algebras*, Vol. I (1971), Vol. II (1985), NH.

Herbrand, J. (1930) Recherches sur la théorie de la démonstration, *Travaux de la Société des Sciences et des Lettres de Varsovie, III*, 33, 33–160. (1971) *Logical Writings*. Harvard University Press and Reidel.

Hermes, H. (1965) *Enumerability, Decidability, Computability*, Springer (second edition, 1969)

Heyting, A. (1956) *Intuitionism*, NH.

Higman, G. (1961) Subgroups of finitely presented groups, *Proc. Royal Soc., Ser. A*, 262, 455–475.

Hilbert, D. and W. Ackermann (1950) *Principles of Mathematical Logic*, Chelsea.

Hilbert, D. and P. Bernays (1934, 1939) *Grundlagen der Mathematik*, Vol. I (1934), Vol. II (1939), Springer (second edition, 1968, 1970).

Hintikka, J. (1955a) Form and content in quantification theory, *Acta Phil. Fennica*, 11–55. (1955b) Notes on the quantification theory, *Comment. Phys.-Math., Soc. Sci. Fennica*, 17, 1–13.

Howard, P.E. (1973) Limitations on the Fraenkel–Mostowski method, *JSL*, 38, 416–422.

Hrbacek, K. and T. Jech (1978) *Introduction to Set Theory*, Academic, (second edition, Dekker, 1984).

Hughes, G.E. and M.J. Creswell (1968) *An Introduction to Modal Logic*, Methuen.

Isbell, J. (1966) Structure of categories, *Bull. AMS*, 72, 619–655.

Jaśkowski, S. (1936) Recherches sur le système de la logique intuitioniste, *Acta Sci. Ind.*, Paris, 393, 58–61.

Jech, T. (1973) *The Axiom of Choice*, NH. (1978). *Set Theory*, Academic.

Jensen, R.B. (1968–69) On the consistency of a slight (?) modification of Quine's New Foundations, *Synthese*, 19, 250–263.

Jeroslow, R.G. (1971) Consistency statements in formal theories, *FM*, 72, 17–40. (1972) On the encodings used in the arithmetization of mathematics, unpublished manuscript. (1973) Redundancies in the Hilbert–Bernays derivability conditions for Gödel's second incompleteness theorem, *JSL*, 38, 359–367. (1976) Consistency statements in formal theories, *FM*, 72, 17–40.

Kalmár, L. (1935) Über die Axiomatisierbarkeit des Aussagenkalküls, *Acta Sci. Math.*, 7, 222–243. (1936) Zuruckführung des Entscheidungsproblems auf den Fall von Formeln mit einer einzigen binären Funktionsvariablen, *Comp. Math.*, 4, 137–144.

Kamke, E. (1950) *Theory of Sets*, Dover.

Kanamori, A. (1994) *The Higher Infinite*, Springer.

Kaye, R. (1991) *Models of Peano Arithmetic*, Oxford University Press, (1994) *Autonorphisms of First Order Structures*, Oxford University Press.

Keisler, H.J. (1976) *Elementary Calculus: An Approach Using Infinitesimals*, Prindle, Weber & Schmidt.

Kelley, J. (1955) *General Topology*, Van Nostrand (Springer, 1975).

Kemeny, J. (1949) Type theory vs. set theory, Ph.D. thesis, Princeton.

Kirby, L. and J. Paris (1977) Initial segments of models of Peano's axioms, *Proc. Bierutowice Conf. 1976*, Lecture Notes in Math., Springer, 211- 226.

Kleene, S.C. (1936a) General recursive functions of natural numbers, *Math. Ann.*, 112, 727–742 (reprinted in Davis, 1965). (1936b) λ-definability and recursiveness, *Duke Math. J.*, 2, 340–353. (1943) Recursive predicates and quantifiers, *Trans. AMS*, 53, 41–73 (reprinted in Davis, 1965). (1952) *Introduction to Metamathematics*, Van Nostrand. (1955a) Hierarchies of number-theoretic predicates, *Bull. AMS*, 61, 193–213. (1955b) Arithmetical predicates and function quantifiers, *Trans. AMS*, 79, 312–340.

Kleene, S.C. and E.L. Post (1954) The upper semi-lattice of degrees of recursive unsolvability, *Ann. Math.*, 59, 379–407.

Kleene, S.C. and R.E. Vesley (1965) *The Foundations of Intuitionist Mathematics*, NH.

Koslow, A. (1992) *A Structuralist Theory of Logic*, Cambridge University Press.

Kreisel, G. and J.-L. Krivine (1967) *Elements of Mathematical Logic*, NH.

Krivine, J.-L. (1971) *Introduction to Axiomatic Set Theory*, Reidel.

Kruse, A.H. (1966) Grothendieck universes and the super-complete models of Shepherdson, *Comp. Math.*, 17, 86–101.

Kunen, K. (1980) *Set Theory. An Introduction to Independence Proofs*, NH.

Kuratowski, K. (1921) Sur la notion d'ordre dans la théorie des ensembles, *FM*, 2, 161–171.

Lambek, J. (1961) How to program an infinite abacus, *Canadian Math. Bull*, 4, 295–302; 5, 297.

Langford, C.H. (1927) Some theorems on deducibility, *Ann. Math.*, I, 28, 16–40; II, 28, 459–471.

Lévy, A. (1960) Axiom schemata of strong infinity, *Pacific J. Math.*, 10, 223–238. (1965) The Fraenkel–Mostowski method for independence proofs in set theory, *The Theory of Models, Proceedings of the 1963 International Symposium at Berkeley*, NH, 221–228. (1978) *Basic Set Theory*, Springer.

Lewis, C.I. and C.H. Langford (1960) *Symbolic Logic.*, Dover (reprint of 1932 edition).

Lewis, H.R. (1979) *Unsolvable Classes of Quantificational Formulas*, Addison-Wesley.

Lindenbaum, A. and A. Mostowski (1938) Über die Unabhängigkeit des Auswahlaxioms und einiger seiner Folgerungen, *Comptes Rendus Sciences Varsovie*, III, 31, 27–32.

Lindström, P. (1969) On extensions of elementary logic, *Theoria*, 35, 1–11.

Löb, M.H. (1955) Solution of a problem of Leon Henkin, *JSL*, 20, 115–118.

Loś, J. (1954a) Sur la théorème de Gödel sur les theories indénombrables, *Bull. de l'Acad. Polon. des Sci.*, III, 2, 319–320. (1954b) On the existence of a linear order in a group, *ibid.*, 21–23. (1954c) On the categoricity in power of elementary deductive systems and some related problems, *Coll. Math.*, 3, 58–62. (1955a) The algebraic treatment of the methodology of elementary deductive systems, *Studia Logica*, 2, 151–212. (1955b) Quelques remarques, théorèmes et problèmes sur les classes definissables d'algèbres, *Math. Interpretations of Formal Systems*, NH, 98–113.

Löwenheim, L. (1915) Ueber Möglichkeiten im Relativkalkül, *Math. Ann.*, 76, 447–470.

Luxemburg, W.A.J. (1962) *Non-Standard Analysis*, Caltech Bookstore, Pasadena. (1969) *Applications of Model Theory to Algebra, Analysis, and Probability*, Holt, Rinehart and Winston. (1973) What is non-standard analysis? *Papers in the Foundations of Mathematics, Amer. Math. Monthly*, 80, No. 6, Part II, 38–67.

Machtey, M., and P. Young (1978) *An Introduction to the General Theory of Algorithms*, NH.

MacLane, S. (1971) Categorical algebra and set-theoretic foundations, *Proc. Symp. Pure Mathematics*, AMS, XIII, Part I, 231–240.

Maclaughlin, T. 1961. A muted variation on a theme of Mendelson, *ZML*, 17, 57–60.

Magari, R. (1975) The diagonalizable algebras, *Boll. Unione Mat. Italiana* (4), 12, 117–125.

Malinowski, G. (1993) *Many-Valued Logics*, Oxford University Press.

Manzano, M. (1996) *Extensions of First Order Logic*, Cambridge University Press.

Margaris, A. (1967) *First-Order Mathematical Logic*, Blaisdell (Dover, 1990).

Markov, A.A. (1954) *The Theory of Algorithms*, Tr. Mat. Inst. Steklov, XLII (translation: Office of Technical Services, U.S. Department of Commerce, 1962).

Matiyasevich, Yu. (1970) Enumerable sets are Diophantine, *Doklady Akad. Nauk SSSR*, 191, 279–282 (English translation, *Soviet Math. Doklady*, 1970, 354–357). (1993) *Hilbert's Tenth Problem*, MIT Press.

McKenzie, R. and R.J. Thompson (1973) An elementary construction of unsolvable word problems in group theory, *Word Problems* (eds W.W. Boone, F.B. Cannonito and R.C. Lyndon), NH.

McKinsey, J.C.C. and A. Tarski (1948) Some theorems about the sentential calculi of Lewis and Heyting, *JSL*, 13, 1–15.

Melzak, Z.A. (1961) An informal arithmetical approach to computability and computation, *Canadian Math. Bull.*, 4, 279–293.

Mendelson, E. (1956a) Some proofs of independence in axiomatic set theory, *JSL*, 21, 291–303. (1956b) The independence of a weak axiom of choice, *ibid.*, 350–366. (1958) The axiom of Fundierung and the axiom of choice, *Arch.*, 4, 65–70. (1961) On non-standard models of number theory, *Essays on the Foundations of Mathematics*, Magnes, Jerusalem, 259–268. (1970) *Introduction to Boolean Algebra and Switching Circuits*, Schaum, McGraw-Hill. (1973) *Number Systems and the Foundations of Analysis*, Academic. (reprint Krieger, 1985). (1990) Second Thoughts about Church's thesis and mathematical proofs, *J. Philosophy*, 225–233.

Meredith, C.A. (1953) Single axioms for the systems (C,N), (C,O) and (A,N) of the two-valued propositional calculus, *J. Comp. Syst.*, 3, 155–164.

Monk, J.D. (1976) *Mathematical Logic*, Springer. (1980) *Introduction to Set Theory*, Krieger.

Montagna, F. (1979) On the diagonalizable algebra of Peano arithmetic, *Boll. Un. Mat. Ital.*, 5, 16-B, 795–812.

Montague, R. (1961a) Semantic closure and non-finite axiomatizability, *Infinitistic Methods*, Pergamon, 45–69. (1961b) Fraenkel's addition to the axioms of Zermelo, *Essays on the Foundations of Mathematics*, Magnes, Jerusalem, 91–114.

Montague, R. and R.L. Vaught (1959) Natural models of set theories, *FM*, 47, 219–242.

Moore, G.H. (1980) Beyond first-order logic: The historical interplay between mathematical logic and set theory, *History and Philosophy of Logic*, 1, 95–137. (1982) *Zermelo's Axiom of Choice: Its Origin, Development and Influence*, Springer. (1988) The emergence of first-order logic, *History and Philosophy of Modern Mathematics* (eds W. Aspray and P. Kitcher), University of Minnesota Press, 95–135.

Morley, M. (1965) Categoricity in power, *Trans. AMS*, 114, 514–538.

Morse, A. (1965) *A Theory of Sets*. Academic.

Mostowski, A. (1939) Ueber die Unabhängigkeit des Wohlordnungsatzes vom Ordnungsprinzip, *FM*, 32, 201–252. (1947) On definable sets of positive integers, *ibid.*, 34, 81–112. (1948) On the principle of dependent choices, *ibid.*, 35, 127–130. (1951a) Some impredicative definitions in the axiomatic set theory, *ibid.*, 37, 111–124 (also 38, 1952, 238). (1951b) On the rules of proof in the pure functional calculus of the first order. *JSL*, 16, 107–111.

Myhill, J. (1955) Creative sets, *ZML*, 1, 97–108.

Nerode, A. (1993) *Logic for Applications*, Springer.

Neumann, J. von (1925) Eine Axiomatisierung der Mengenlehre, *J. für Math.*, 154, 219–240 (also 155, 128) (English translation in Van Heijenoort, 1967, 393–413). (1928) Die Axiomatisierung der Mengenlehre, *Math. Z.*, 27, 669–752.

Nicod, J.G. (1917) A reduction in the number of primitive propositions of logic, *Proc. Camb. Phil. Soc.*, 19, 32–41.

Novak, I.L. (Gal, L.N.) (1951) A construction for models of consistent systems, *FM*, 37, 87–110.

Novikov, P.S. (1955) On the algorithmic unsolvability of the word problem for group theory, *Tr. Mat. Inst. Steklov*, 44 (*Amer. Math. Soc. Translations, Series* 2, 9, 1–124)

Oberschelp, A. (1991) On pairs and tuples. *ZML*, 37, 55–56.
Oberschelp, W. (1958) Varianten von Turingmaschinen, *Arch.*, 4, 53–62.
Orey, S. (1956a) On ω-consistency and related properties, *JSL*, 21, 246–252. (1956b) On the relative consistency of set theory, *ibid.*, 280–290.
Parikh, R. (1971) Existence and feasibility in arithmetic, *JSL*, 36, 494–508.
Paris, J.B. (1972) On models of arithmetic, *Conference in Math. Logic – London, 1970*, Springer, 251–280. (1978) Some independence results for Peano arithmetic, *JSL*, 43, 725–731.
Paris, J. and L. Harrington (1977) A mathematical incompleteness in Peano arithmetic, *HML*, 1133–1142.
Peano, G. (1891) Sul concetto di numero, *Rivista di Mat.*, 1, 87–102.
Péter , R. (1935) Konstruktion nichtrekursiver Funktionen, *Math. Ann.*, 111, 42–60. (1967) *Recursive Functions*, Academic.
Pincus, D. (1972) ZF consistency results by Fraenkel–Mostowski methods, *JSL*, 37, 721–743.
Post, E.L. (1921) Introduction to a general theory of elementary propositions, *Am. J. Math.*, 43, 163–185. (1936) Finite combinatory process-formulation 1, *JSL*, 1, 103–105 (reprinted in Davis, 1965). (1941) *The Tw-Valued Iterative Systems of Mathematical Logic*, Princeton University Press. (1943) Formal reductions of the general combinatorial decision problem, *Am. J. Math.*, 65, 197–215. (1944) Recursively enumerable sets of positive integers and their decision problems, *Bull. AMS*, 50, 284–316 (reprinted in Davis. 1965). (1947) Recursive unsolvability of a problem of Thue, *JSL*, 12, 1–11 (reprinted in Davis, 1965). (1994) *Solvability, Provability, Definability: The Collected Works of Emil L. Post* (ed. M. Davis), Birkhäuser.
Presburger, M. (1929) Ueber die Vollständigkeit eines gewissen Systems der Arithmetik ganzer Zahlen in welchem die Addition als einziger Operation hervortritt, *Comptes Rendus, 1 Congrès des Math. des Pays Slaves*, Warsaw, 192–201, 395.
Putnam, H. (1957) Decidability and essential undecidability, *JSL*, 22, 39–54.
Quine, W.V. (1937) New foundations for mathematical logic, *Am. Math. Monthly*, 44, 70–80. (1951) *Mathematical Logic* (second revised edition), Harvard University Press. (1954) Quantification and the empty domain, *JSL*, 19, 177–179 (reprinted in Quine, 1965, 220–223). (1963) *Set Theory and its Logic*, Harvard University Press. (revised edition, 1969) (1965) *Selected Logical Papers*, Random House.
Rabin, M. (1958) On recursively enumerable and arithmetic models of set theory, *JSL*, 23, 408–416. (1977) Decidable theories, *HML*, 595–629.
Ramsey, F.P. (1925) New foundations of mathematics, *Proc. London Math. Soc.*, 25, 338–384.
Rasiowa, H. (1956) On the ε-theorems, *FM*, 43, 156–165. (1974) *An Algebraic Approach to Non-Classical Logics*, NH.
Rasiowa, H. and R. Sikorski (1951) A proof of the completeness theorem of Gödel, *FM*, 37, 193–200. (1952) A proof of the Skolem–Löwenheim theorem, *ibid.*, 38, 230–232. (1963) *The Mathematics of Metamathematics*, Państwowe Wydawnictwo Naukowe.
Rescher, N. (1969) *Many-Valued Logics*, McGraw-Hill.
Rice, H.G. (1953) Classes of recursively enumerable sets and their decision problems, *Trans. AMS*, 74, 358–366.

Robinson, A. (1951) *On the Metamathematics of Algebra*, NH. (1952) On the application of symbolic logic to algebra, *Int. Cong. Math., Cambridge*, 1, 686–694. (1966) *Non-Standard Analysis*, NH. (1979) *Selected Papers*, Vol. 1, *Model Theory and Algebra*, NH; Vol. 2, *Nonstandard Analysis and Philosophy*, Yale University Press.

Robinson, J.A. (1965) A machine-oriented logic based on the resolution principle, *J. Assoc. Comp. Mach.*, 12, 23–41.

Robinson, Julia. (1949) Definability and decision problems in arithmetic, *JSL*, 14, 98–114. (1950) General recursive functions, *Proc. AMS*, 1, 703–718. (1952) Existential definability in arithmetic, *Trans. AMS*, 72, 437–449.

Robinson, R.M. (1937) The theory of classes. A modification of von Neumann's system, *JSL*, 2, 69–72. (1947) Primitive recursive functions, *Bull. AMS*, 53, 925–942. (1948) Recursion and double recursion, *Bull. AMS*, 54, 987–993. (1950) An essentially undecidable axiom system, *Proc. Int. Cong. Math., Cambridge*, 1, 729–730.

Rogers, H., Jr (1959) Computing degrees of unsolvability, *Math. Ann.*, 138, 125–140. (1967) *Theory of Recursive Functions and Effective Computability*, McGraw-Hill.

Rosenbloom, P. (1950) *Elements of Mathematical Logic*, Dover.

Rosser, J.B. (1936) Extensions of some theorems of Gödel and Church, *JSL*, 87–91 (reprinted in Davis, 1965). (1937) Gödel theorems for non-constructive logics, *JSL*, 2, 129–137. (1939) An informal exposition of proofs of Gödel's theorem and Church's theorem, *JSL*, 53–60 (reprinted in Davis, 1965). (1942) The Burali-Forti paradox, *JSL*, 7, 1–17. (1953) *Logic for Mathematicians*, McGraw-Hill (second edition, Chelsea, 1978). (1954) The relative strength of Zermelo's set theory and Quine's New Foundations, *Proc. Int. Cong. Math., Amsterdam*, III, 289–294. (1969) *Simplified Independence Proofs*, Academic.

Rosser, J.B., and A. Turquette (1952) *Many-Valued Logics*, NH (second edition, 1977, Greenwood).

Rosser, J.B., and H. Wang (1950) Non-standard models for formal logics, *JSL*, 15, 113–129.

Rotman, J.J. (1973) *The Theory of Groups*. Allyn & Bacon (second edition).

Rubin, H. and J. Rubin (1963) *Equivalents of the Axiom of Choice*, NH.

Rubin, J. (1967) *Set Theory for the Mathematician*, Holden-Day.

Russell, B. (1908) Mathematical logic as based on the theory of types, *Am. J. Math.*, 30, 222–262.

Ryll-Nardzewski, C. (1953) The role of the axiom of induction in elementary arithmetic, *FM*, 39, 239–263.

Sambin, G. (1976) An effective fixed point theorem in intuitionistic diagonalizable algebras, *Studia Logica*, 35, 345–361.

Schütte, K. (1951). Beweistheoretische Erfassung der unendlichen Induktion in der Zahlentheorie, *Math. Ann.*, 122, 368–389.

Shannon, C. (1938) A symbolic analysis of relay and switching circuits, *Trans. Amer. Inst. Elect. Eng.*, 57, 713–723.

Shapiro, S. (1988) The Lindenbaum construction and decidability, *NDJFL*, 29, 208–213. (1991) *Foundations without Foundationalism: A Case for Second-order Logic*, Oxford University Press.

Shepherdson, J.C. (1951–53) Inner models for set theory, *JSL*, 1, 16, 161–190; II, 17, 225–237; III, 18, 145–167. (1961) Representability of recursively enumerable sets in formal theories, *Arch.*, 5, 119–127.

Shepherdson, J.C. and H.E. Sturgis (1963) Computability of recursive functions, *J. Assoc. Comp. Mach.*, 10, 217–255.

Shoenfield, J. (1954) A relative consistency proof, *JSL*, 19, 21–28. (1955) The independence of the axiom of choice, Abstract, *ibid.*, 20, 202. (1961) Undecidable and creative theories, *FM*, 49, 171–179. (1967) *Mathematical Logic*. Addison-Wesley. (1971) Unramified forcing, *Proc. Symp. Pure Math.*, 13, AMS, 357–381.

Sierpinski, W. (1947) L'Hypothèse généralisée du continu et l'axiome du choix, *FM*, 34, 1–5. (1958) *Cardinal and Ordinal Numbers*, Warsaw.

Sikorski, R. (1960) *Boolean Algebra*, Springer (third edition, 1969).

Skolem, T. (1919) Untersuchungen über die Axiome des Klassenkalküls und über Produktations- und Summationsprobleme, welche gewisse Klassen von Aussagen betreffen. *Skrifter-Vidensk*, Kristiana, I, 1–37. (1920) Logisch-kombinatorische Untersuchungen über die Erfüllbarkeit oder Beweisbarkeit mathematischer Sätze, *ibid.*, 1–36 (English translation in Van Heijenoort, 1967, 252–263). (1923) Einige Bemerkungen zur axiomatischen Begründung der Mengenlehre, *Wiss. Vorträge gehalten auf dem 5. Kongress der skandinav. Mathematiker in Helsingförs*, 1922, 217–232 (reprinted in Van Heijenoort, 1967, 290–301). (1934) Ueber die Nicht-Characterisierbarkeit der Zahlenreihe mittels endlich oder abzählbar unendlich vieler Aussagen mit ausschliesslich Zahlenvariablen, *FM*, 23, 150–161.

Smoryński, C. (1977) The incompleteness theorems, *HML*, 821–866. (1981) Fifty years of self-reference, *NDJFL*, 22, 357–374. (1985) *Self-Reference and Modal Logic*, Springer. (1991) *Logical Number Theory I*, Springer.

Smullyan, R. (1961) *Theory of Formal Systems*, Princeton University Press. (1968) *First-Order Logic*, Springer (reprint Dover, 1995). (1978) *What is the Name of This Book*, Prentice Hall. (1985) *To Mock a Mockingbird*, Knopf. (1992) *Gödel's Incompleteness Theorems*, Oxford University Press. (1993) *Recursion Theory for Metamathematics*, Oxford University Press. (1994) *Diagonalization and Self-Reference*, Oxford University Press.

Solovay, R.M. (1976) Provability interpretations of modal logic, *Israel J. Math.*, 25, 287–304.

Sonner, J. (1962) On the formal definition of categories, *Math. Z.*, 80, 163–176.

Specker, E. (1949) Nicht-konstruktiv beweisbare Sätze der Analysis, *JSL*, 14, 145–148. (1953) The axiom of choice in Quine's 'New Foundations for Mathematical Logic', *Proc. Natl. Acad. Sci. USA*, 39, 972–975. (1954) Verallgemeinerte Kontinuumhypothese und Auswahlaxiom, *Archiv der Math.*, 5, 332 337. (1957) Zur Axiomatik der Mengenlehre (Fundierungs- und Auswahlaxiom), *ZML*, 3, 173–210. (1958) Dualität, *Dialectica*, 12, 451–465. (1962) Typical ambiguity, *Logic, Methodology and Philosophy of Science, Proc. Int. Cong., 1960*, Stanford, 116–124.

Stone, M. (1936) The representation theorem for Boolean algebras, *Trans. AMS*, 40, 37–111.

Stroyan, K.D. and W.A.J. Luxemburg (1976) *Introduction to the Theory of Infinitesimals*, Academic.

Suppes, P. (1960) *Axiomatic Set Theory*, Van Nostrand (Dover, 1972).

Szmielew, W. (1955) Elementary properties of abelian groups, *FM*, 41, 203–271.

Takeuti, G. and W.M. Zaring (1971) *Introduction to Axiomatic Set Theory*, Springer. (1973) *Axiomatic Set Theory*, Springer.

Tarski, A. (1923) Sur quelques théorèmes qui equivalent à l'axiome de choix, *FM*, 5, 147–154. (1925) Sur les ensembles finis, *ibid.*, 6, 45–95. (1933) Einige Berach-

tungen über die Begriffe der ω-Widerspruchsfreiheit und der ω-Vollständigkeit, *Monatsh. Math. Phys.*, 40, 97–112. (1936) Der Wahrheitsbegriff in den formalisierten Sprachen, *Studia Philos.*, 1, 261–405 (English translation in Tarski, 1956). (1938) Ueber unerreichbare Kardinalzahlen, *FM*, 30, 68–89. (1944) The semantic conception of truth and the foundations of semantics, *Philos. and Phenom. Res.*, 4, 341–376. (1951) *A Decision Method for Elementary Algebra and Geometry*, Berkeley. (1952) Some notions and methods on the borderline of algebra and metamathematics, *Int. Cong. Math., Cambridge, Mass, 1950*, AMS, 705–720. (1956) *Logic, Semantics, Metamathematics*, Oxford University Press. (second edition, 1983, J. Corcoran (ed.), Hackett).

Tarski, A., A. Mostowski and R. Robinson (1953) *Undecidable Theories*, NH.

Tarski, A. and R.L. Vaught (1957) Arithmetical extensions of relational systems, *Comp. Math.*, 18, 81–102.

Troelstra, A.S. (1969) *Principles of Intuitionism*, Springer.

Turing, A. (1936–37) On computable numbers, with an application to the Entscheidungsproblem, *Proc. London Math. Soc.*, 42, 230–265; 43, 544–546. (1937) Computability and λ-definability, *JSL*, 2, 153–163. (1948) Practical forms of type theory, *ibid.*, 13, 80–94. (1950a) The word problem in semigroups with cancellation, *Ann. Math.*, 52, 491–505 (review by W.W. Boone, *ibid.*, 1952, 74–76). (1950b) Computing machinery and intelligence, *Mind*, 59, 433–460.

Ulam, S. (1930) Zur Masstheorie in der allgemeinen Mengenlehre, *FM*, 16, 140–150.

Van Benthem, J. and K. Doets. (1983) Higher-order logic. *HPL*, I, 275–330.

Vaught, R.L. (1954) Applications of the Löwenheim–Skolem–Tarski theorem to problems of completeness and decidability, *Indag. Math.*, 16, 467–472.

Waerden, B.L. van der (1949) *Modern Algebra*, Ungar.

Wajsberg, M. (1933) Untersuchungen über den Funktionenkalkül für endliche Individuenbereiche, *Math. Annalen*, 108, 218–228.

Wang, H. (1949) On Zermelo's and von Neumann's axioms for set theory, *Proc. Natl. Acad. Sci. USA*, 35, 150–155. (1950) A formal system of logic, *JSL*, 15, 25–32. (1957) The axiomatization of arithmetic, *ibid.*, 22, 145–158.

Whitehead, A.N., and B. Russell (1910–13) *Principia Mathematica*, Vols I–III, Cambridge University Press.

Wiener, N. (1914) A simplification of the logic of relations, *Proceedings of the Cambridge Philosophical Society*, 17, 387–390 (reprinted in Van Heijenoort, 1967, 224–227).

Yasuhara, A. (1971) *Recursive Function Theory and Logic*, Academic.

Zeeman, E.C. (1955) On direct sums of free cycles, *J. London Math. Soc.*, 30, 195–212.

Zermelo, E. (1904) Beweis, dass jede Menge wohlgeordnet werden kann, *Math. Annalen*, 59, 514–516 (English translation in Van Heijenoort, 1967, 139–141). (1908) Untersuchungen über die Grundlagen der Mengenlehre, *ibid.*, 65, 261–281 (English translation in Van Heijenoort, 1967, pp. 199–215).

Zuckerman, M. (1974) *Sets and Transfinite Numbers*, Macmillan.

Notation

Index