MATHEMATICAL LOGIC

by

R. L. GOODSTEIN, Ph.D., D.Lit.

Professor of Mathematics in the
University of Leicester

LEICESTER UNIVERSITY PRESS
1957

Printed by Sir Isaac Pitman & Sons, Ltd.
at the Pitman Press, Bath, England
for the
Leicester University Press

PREFACE

THE aim of this little book is to introduce teachers of mathematics to some of the remarkable results which have been obtained in mathematical logic during the past twenty-five years. The book is designed to be read by mathematicians who have little or no previous knowledge of symbolic logic, and is largely self-contained in the sense that the proofs of major results are given in detail. A great many different facets of the subject have been briefly sketched, but rigour has not been intentionally sacrificed for ease of reading, nor has generality been pursued for its own sake.

A work of this kind is necessarily indebted to all the leading thinkers in its field and some acknowledgement of this indebtedness is made in the Bibliographical Notes at the end of the book. To Mr. John Hooley my warmest thanks are due for preparing the index and for generous help in correcting the proofs. I am again indebted to the compositors and printers of the Pitman Press for their skill and attention to detail.

R. L. GOODSTEIN

UNIVERSITY COLLEGE
LEICESTER
February, 1956

CONTENTS

PAGE

INTRODUCTION

THE FUNCTION OF MATHEMATICAL LOGIC 1

The definition of number and variable

CHAPTER I

THE SENTENCE CALCULUS 11

Truth tables. Arithmetical decision procedure. Three-valued logic. Axiomatic theory. Completeness and freedom from contradiction. Independence of the axioms. Intuitionistic logic. Polish bracket-free notation. Natural inference.

CHAPTER II

PREDICATE CALCULUS 28

Axiomatic theory. Freedom from contradiction. The deduction theorem. Natural inference. Validity and satisfiability. Decision procedure for monadic predicates. Gödel's completeness theorem.

CHAPTER III

NUMBER THEORY 44

The systems Z and Z_r. Primitive recursive functions. Ordinal recursions. The calculus of λ-conversion. Recursive arithmetic. Arithmetical relations.

CHAPTER IV

THE INCOMPLETENESS OF ARITHMETIC 73

Gödel numbering, and the arithmetization of syntax. Undecidable statements. Impossibility of characterizing the natural numbers by an axiomatic system. The decision problem. The undecidability of arithmetic and the undecidability of predicate logic.

CHAPTER V PAGE

EXTENDED PREDICATE LOGIC 89

Class logic. Stratification. Descriptions operator. Ordered pairs. Class of natural numbers. The relativity of class concepts.

NOTES AND BIBLIOGRAPHY 98

INTRODUCTION

THE FUNCTION OF MATHEMATICAL LOGIC

The Definition of Number and Variable

THE function of mathematical logic is to reveal and codify the logical processes employed in mathematical reasoning and to clarify the concepts of mathematics; it is itself a branch of mathematics, employing mathematical symbolism and technique, a branch which has developed in its entirety during the past hundred years and which in its vigour and fecundity and the power and importance of its discoveries may well claim to be in the forefront of modern mathematics.

Because of the part which it plays in elucidating the nature of mathematics, mathematical logic has a special importance for the teacher of mathematics; current discoveries in almost every branch of mathematics throw light on some elementary topic, revealing unsuspected connections or limitations, but the discoveries which have been made in mathematical logic illuminate not simply particular topics but almost every facet of mathematics.

The young teacher, fresh from his Honours Degree course in mathematics, soon finds that he has added nothing to the knowledge of elementary arithmetic which he gained at school himself. He knows no better now than then why, to divide, you turn a fraction upside down and multiply, but if his University course was a good one he should be in a position to find the reason out for himself; he should be capable of asking himself just what division and multiplication mean and how they are related, and he will recall to mind some comparable situation, perhaps his initiation into the mysteries of complex numbers, or quaternions. But if his pupils ask him what is a number or a variable, a line or an angle; how a proof proves anything, whether truth is proof, and whether we can be sure that only true theorems are provable, then, unless his training was exceptional, he will neither know the answers nor be in a position to think them out for himself. Of course the answers which modern mathematical logic gives to these questions are answers for the teacher and not his pupils, the extent to which

they can be communicated depending both upon the teacher's skill and his pupils' maturity, but the important thing is that the answers are now available even though they have, in some instances, only very recently been discovered.

The concept of number

As all mathematics depends upon the concept of a natural number the task of clarifying this concept continues to be, as it always has been, one of mathematical logic's primary concerns. The answer to the question, 'What is a number?', will, however, be seen to depend, in part at any rate, upon the answer to the wider question, 'What is logic?'. We shall see that there are different levels of logic admitting different definitions of number, and this relativity of concepts is itself a discovery of fundamental importance.

The intuitive idea of number is that it is a property of collections; the number two, for instance, being what all pairs have in common. If, however, number is a property of a collection it is not a property in the same sense in which the colour of a door is a property of the door. A door remains a door whatever colour it is painted, and it makes perfectly good sense to say that a door which was formerly white and is now green is the same door, or even that a door has *no* colour (for instance, if it is made of glass) but it makes no sense at all to say of some bundle of candles that it has no number. Number is intrinsic to a collection, an essential characteristic, a *sine qua non*, not a physical property like temperature, colour, or weight. This is not to say, however, that number is a mental quality, something in the mind of the observer alone; number is a logical, not a psychological concept.

The Frege-Russell definition

The first purely logical definition of number was published by the German mathematician Gottlob Frege in 1894, but Frege's definition remained virtually unknown until it was rediscovered by Bertrand Russell in 1903. The Frege-Russell definition bases number on the notion of *class similarity*. Two classes, or collections, are said to be similar (or homographic) if they are one-to-one related, that is, if a correspondence can be established between the two collections which relates to

each member of one collection a unique mate in the other and to each member of the second collection a unique mate in the first. If similarity is to serve in the definition of number it is essential, of course, that the definition of similarity itself should be entirely free from the number concept, and so we must scrutinize the definition to make sure that such notions as 'each' and 'unique' which it contains are not themselves derived from the concept of number. In the first place the definition asserts the existence of a certain relationship between two collections. In making the definition we are therefore presupposing two things of our logic; that it is capable of expressing relationship and asserting existence. In the system of logic which originated in Russell and Whitehead's *Principia Mathematica* relationship is identified with the class of couples having the relationship in question, class itself being taken as a primitive notion. The definition of similarity, however, not only asserts the existence of a relationship but affirms this relationship of *every* element of the classes concerned; for here we are using the term 'each' in the sense of 'everyone'. Thus the logic we use must not only be capable of asserting existence but also universality; we must be able to say of a collection that *all* its members have some property. Accordingly, if we denote the class membership relation by ε (the initial letter of the Greek word for *is*) the relation of identity by $=\!=$ and similarity by $\sim$, writing, for any relation R, $a\mathrm{R}b$ to express the fact that a stands in the relation R to b, then the definition of similarity of two classes A and B takes the form: $\mathrm{A} \sim \mathrm{B}$ if, and only if, there is relation R such that, for every x, if $x \varepsilon \mathrm{A}$, then there is a y such that $x\mathrm{R}y$ and $y \varepsilon \mathrm{B}$, and for every y, if $y \varepsilon \mathrm{B}$, then there is an x such that $x\mathrm{R}y$ and $x \varepsilon \mathrm{A}$, and further, for every x, y, and z, if $x\mathrm{R}y$ and $x\mathrm{R}z$, then $y =\!= z$, and if $x\mathrm{R}z$ and $y\mathrm{R}z$ then $x =\!= y$.

In this formulation the first and second implications state that each member of one class has a mate in the other, and the next two that these mates are unique. The *structure* of the definition and its purely logical character is more easily seen if we introduce some abbreviations. To assert existence we use the symbol $\exists$, the initial letter of the word existence, reversed, and read 'there is a' and for universality $\forall$ (the initial letter of the word all, upside down) and read 'for all'; for an implication

'if . . . then —' we write '. . . → —' and read '. . . implies —', and for conjunction the familiar ampersand. The double implication, '. . . if, and only if —' is denoted by '. . . ↔ —'. With these abbreviations, and a judicious use of brackets, the definition runs:

$$(\Lambda \sim \mathrm{B}) \leftrightarrow (\exists \mathrm{R})\{(\forall x)[(x \varepsilon \mathrm{A}) \rightarrow (\exists y)(x\mathrm{R}y \ \& \ (y \varepsilon \mathrm{B}))] \ \& \ (\forall y)[(y \varepsilon \mathrm{B}) \rightarrow (\exists x)(x\mathrm{R}y \ \& \ x \varepsilon \mathrm{A})] \ \& \ (\forall x)(\forall y)(\forall z)[(x\mathrm{R}y \ \& \ x\mathrm{R}z \rightarrow (y = z)) \ \& \ (x\mathrm{R}z \ \& \ y\mathrm{R}z \rightarrow (x = y))]\}$$

The 'bricks' of which it is built are (signs for) conjunction and implication, the universal and existential *quantifiers* $\forall$ and $\exists$, *variables* x, y, z, for class members, a relation variable R, three definite relation signs, ε, $\sim$, and $=$, and finally two class variables A and B. Apart from the class variables A, B, each of the class member variables (usually called the object variables), and the relation variable, come within the scope of (i.e. are the subjects of) a quantifier and are said to be *bound* by the quantifier within whose scope they come. The class variables A and B on the contrary are not subject to any quantifier and are said to be *free*. We are not yet in a position to make a complete analysis of the concept of a variable but some preliminary remarks may conveniently be made at this stage. To avoid confusion obviously some conventions are needed about the letters used for the several categories of variables; it is not possible for these conventions to be more than a passing convenience since the categories which we have so far introduced are neither exhaustive nor necessarily independent of one another. For the present then we shall use small Roman letters as object variables, and to ensure an unlimited supply we shall construct new letters, when needed, by affixing dashes or stars, thus x^* or x'''; for relation variables we shall use the Roman capitals R and S, and for classes the capitals A, B, C, the number of letters being increased when needed by affixing dashes or stars. If $\mathscr{A}(x)$ (read 'x has the property of $\mathscr{A}$') denotes an expression in which 'x' appears free and if the statement $\mathscr{A}(x)$ does not contain 'y', then clearly both the statements $(\exists x).\mathscr{A}(x)$, $(\exists y).\mathscr{A}(y)$ say the same thing, namely that there is an object with the property $\mathscr{A}$, which shows that bound variables are dummy letters in the sense that any one

may be totally replaced by another of the same category without changing the sense of the statement in which the replacement is made. A familiar example of a bound variable in elementary mathematics is the letter x in the definite integral $\int_a^b f(x)dx$, where, provided that 'a' and 'b' do not contain 'x', we may replace x by any other letter denoting a variable of the same category. As in the case of variables bound by a quantifier the right of substitution in the integral must be limited by certain safeguards. For instance we must not substitute 'f' or 'a' or 'b' for 'x' in the integral. A bound variable is, in principle, totally eliminable. We could for instance employ a functional notation which omitted the argument, and write the integral simply as $\mathrm{I}_f(a, b)$ or as $\mathrm{I}(f; a,b)$ if we want to stress the fact that the function variable f is in fact an argument in the integral, or we could simply leave a blank space, writing $\int_a^b f(\ )d(\ )$. Where more than one bound variable is simultaneously involved some device would be needed to correlate the appropriate blanks; for instance the partial derivatives $\partial^3 f(x,y)/\partial^2 x\partial y$, $\partial^3 f(y,x)/\partial^2 x\partial y$ are not in general equal and so they could not both be written in the form $\partial^3 f(\ ,\ )/\partial^2(\ )\partial(\ )$. In sufficiently simple cases the correlation could of course be effected by rank alone, but this is not always possible, as the example

$$(\exists x)(\exists y)\{\mathrm{R}(x,y)\ \&\ \mathrm{S}(y,x)\}$$

shows. The desired correlation could, however, be effected by using special brackets, or by means of braces, thus

$$(\exists(\overline{\ }))(\exists(\underline{\ }))\{\mathrm{R}(\overline{\ },\underline{\ })\ \&\ \mathrm{S}(\underline{\ },\overline{\ })\}$$

All this goes to show that bound variables, though dispensable, are a very great convenience and more than that cannot be expected of any notation.

Free variables, like bound variables, are also replaceable by other variables of the same category but they have the additional property of being replaceable by *particulars* of the same category. For instance for the free class variables in the definition of similar classes we may substitute any particular classes we

Counting

The adequacy of the Frege-Russell definition of number is shown by the fact that, for any property $\mathscr{A}$, each of the following statements is provable:

$$\text{not } (\exists x)\mathscr{A}(x) \leftrightarrow (\hat{x}\mathscr{A}(x)\varepsilon 0),$$
$$(\exists x)[\mathscr{A}(x) \ \&\ (\forall y)(\mathscr{A}(y) \rightarrow y \equiv x)] \leftrightarrow \hat{x}\mathscr{A}(x)\varepsilon 1,$$
$$(\exists x)(\exists y)\{\mathscr{A}(x) \ \&\ \mathscr{A}(y) \ \&\ x \neq y \ \&\ (\forall z)[\mathscr{A}(z) \rightarrow (x \equiv z) \text{ or } (y \equiv z)]\} \leftrightarrow \hat{x}\mathscr{A}(x)\varepsilon 2,$$

and so on. The second of these equivalences, for example, says that the class of x's for which $\mathscr{A}(x)$ is true is similar to the class whose only member is 0, if and only if there is an x for which $\mathscr{A}(x)$ is true, and $\mathscr{A}(y)$ is true only when $y \equiv x$. The equivalence is established by showing that the relation R whose sole member is the couple consisting of the unique x for which $\mathscr{A}(x)$ holds and the null class (the unique member of the class 1) establishes a one-one correlation between the classes $\hat{x}\mathscr{A}(x)$ and 1.

As we have already observed the Frege-Russell definition can only be formulated in a logic which is rich enough to assert the existence of classes. Such a logic (a class logic or an extended predicate logic as it is often called) is one in which, for a variety of reasons we shall later discuss, we cannot have complete confidence. Suffice it for the present to say that both Frege's original formulation of class logic as also the more sophisticated later version of the American logician Willard van Orman Quine, were found to contain contradictions on the very eve of publication.

The numerals

A seemingly entirely different approach to the problem of defining number focuses attention, not upon numbers themselves but upon numerals. It is of course evident that numbers are not the same things as numerals, since they have different properties; numbers are even or odd, prime or composite, and numerals are Roman or Arabic, written in ink or cast in iron. The numerals are in fact not numbers but *number signs*; we say *number signs* rather than *signs for numbers* to stress the important point that we do not need to know anything about

please. Thus we may substitute for A the class N of natural numbers and for B the class F of squares (which are similar) or for A the class I of irrational numbers and for B the class A of algebraic numbers (which are not similar). We cannot, however, substitute particulars for bound variables as is evident from the example 'the derivative of x^2 with respect to x is $2x$', in which the substitution of 2 for x produces the nonsense (not entirely unfamiliar to those who have taught a first course in Calculus) 'the derivative of 4 with respect to 2 is 4'.

The free variables in the definition of similarity can readily be replaced by bound variables, simply by prefixing the quantifiers, $(\forall A)(\forall B)$, to the definition, provided that (and this provision is an important one) we permit the application of the quantifiers also to this category of variables. Since, however, we cannot substitute particulars for bound variables the application of the definition formulated in bound variables to particular classes is possible only if we introduce a *derivation rule* enabling us to pass from the universal to the particular. It suffices to permit the inference of $\mathscr{A}(. \ . \ .)$, with a particular in the argument place, from the universal statement $(\forall\xi)\mathscr{A}(\xi)$ where ξ stands for a variable of the same category as the particular.

Conversely, in some contexts, we may replace a variable bound by a universal quantifier by a free variable, since the right we have to substitute *any* particular of the same category for a free variable, shows that an expression '$\mathscr{A}(\xi)$' with a free variable ξ is equivalent to the statement 'any particular has the property $\mathscr{A}$' and this of course is equivalent to 'every particular has the property', which is expressed by '$\forall\xi\mathscr{A}(\xi)$' with a bound variable. That this replacement cannot always be effected is, however, evident from the fact that, for instance, the statements

$$(\forall x)\mathscr{A}(x) \rightarrow (\forall x)\mathscr{A}(x), \ \mathscr{A}(x) \rightarrow (\forall x)\mathscr{A}(x)$$

(i.e. if *every* x has the property $\mathscr{A}$ then every x has the property $\mathscr{A}$, if *any* x has the property $\mathscr{A}$ then every x has the property $\mathscr{A}$) have different meanings, since the first is true for any property $\mathscr{A}$ and the second is false when $\mathscr{A}$ is, for instance, the property of being a prime number.

The cardinal numbers

The next step in the Frege-Russell definition of number is the definition of the *cardinal number of a class*. Denoting by $\hat{\xi}\mathscr{A}(\xi)$ the class of all objects which have the property $\mathscr{A}$ then *the cardinal number of a class* α is defined to be the class $\hat{\xi}(\xi \sim \alpha)$, that is, the class of all classes similar to α; it follows that two similar classes have the same cardinal number. Then, starting with the empty (null) class Λ, which is defined as $\hat{\xi}(\xi \neq \xi)$, the class of all things not identical with themselves (and so a class without members since any object is identical with itself) the number 0 is defined to be the cardinal of the class Λ (so that 0 is a *class*, viz. the class of all classes similar to the null class) and the number 1 is defined as the cardinal of 0 (so that 1 is $\hat{\xi}(\xi \sim 0)$). According to these definitions a class has 0 members if it belongs to 0, and is therefore similar to the null class, and a class has 1 member if it belongs to 1 and is therefore similar to 0, the class whose only member is the null class, which shows that the Frege-Russell definition, though at first sight rather mysterious and peculiar, in fact gives zero and unity their familiar properties. If instead of speaking entirely in terms of classes, we take the membership property of a class instead of the class itself (as in fact Frege himself did) the definition loses some of its strangeness. The cardinal of a class is now described in terms of the property of being similar to the class; the number zero is the property of being similar to the null class and the number one is the property of being similar to the class whose membership condition is the property of having 0 terms.

We could of course proceed now to define in turn the number 2 as the cardinal of the class (0, 1) whose members are 0 and 1, the number 3 as the cardinal of the class (0, 1, 2), and so on, but it is simpler (and more general) to define first the operation '+ 1'. Given any class A of cardinal a we define $a + 1$ to be the cardinal of the class

$$\hat{\xi}(\xi \varepsilon \mathrm{A} \text{ or } \xi \varepsilon \{\mathrm{A}\})$$

where {A} denotes the class whose sole member is A itself. Then 2 is defined as $1 + 1$, 3 as $2 + 1$ and so on, so that the numbers 1, 2, 3 and so on are in fact the cardinals of the classes

$$\{0\}, \{0, \{0\}\}, \{0, \{0\}, \{\{0\}\}\}, \text{ and so on,}$$

where 0 is the cardinal of the null class.

numbers to be able to give an account of the numerals. Numerals are simply signs of a certain category, like money signs, £ and $ or music signs 𝄞 and ♩♫. More exactly the numerals are words spelt in a very primitive alphabet. To illustrate this more readily we shall introduce a notation for numerals which has proved very convenient in studies on the foundations of arithmetic; in this notation the numerals are o, So, SSo, SSSo, and so on, so that we may say that the numerals are spelt with S's and a final o. The more familiar numerals, 1, 2, 3, 4, and so on, are abbreviations, '1' standing for 'So', '2' standing for 'SSo', '3' for 'SSSo', '4' for 'SSSSo', and so on. We shall see that it is quite easy to give an account of arithmetic in terms of numerals in a far more economical logic than is needed to formulate the Frege-Russell numbers.

In this account of arithmetic we shall find that the concept of a number is entirely dispensable, though dispensable only at the cost of considerable inconvenience and lengthy circumlocutions. Does it then follow that the numerals are themselves the subject of arithmetic and not the natural numbers? Whether the answer to this question is in the affirmative or not depends upon what one means by 'the subject of arithmetic'. It is often helpful in seeking to answer a question of this kind to change the setting in which the question is posed, and we shall turn, as one so often does in logical enquiries, to the game of chess. Let us ask, instead of our question about numbers and arithmetic, the exactly parallel question: 'Are the pieces on the chess-board the subject of the rules of chess?'. For instance, is the rule that the king moves only one square at a time, except in castling, a rule about the king of chess or about a piece of wood of particular design? A preliminary answer is that the rule *cannot* be about a particular piece of wood because we might lose it and yet play chess with a lump of sugar in its place. So too the rule that 6 is an even number cannot be a rule about this particular mark in ink, since we could express the rule just as well by writing that 3 + 3 is an even number. None the less, given a chess board and pieces we can teach someone to play chess, using the familiar terms king, queen, pawn, etc., simply as object names (or not using them at all). But if the king of chess is not a particular piece on a particular chess board, what is he? If the number two is not the numeral '2' what is it? To

put the question another way, what is it that makes a particular piece in a particular game of chess, the king piece? It is not the shape of the piece (for we might interchange the king and queen), nor the position of the piece in the game (which may be anywhere on the board). No, what constitutes a piece king are *the moves which it makes*. Thus we may say that the king of chess is one of the *rôles* which a piece plays in a game of chess—the part which the piece plays, not the piece itself. And so too the numbers are the several parts which the numerals play in language. The rules of arithmetic, like the rules of chess, are formulated in terms of the permitted transformations of the number signs. For instance, the rule that the sum of two and three is five is a formulation in terms of *rôles* of the fact that the *formula* '$2 + 3 = 5$' is provable in arithmetic. And if we exchange the parts which the numerals 2 and 5 play, so that each plays the other's part, then the formula '$5 + 3 = 2$' is provable, and is still an expression of the rule that the sum of two and three is five; the formulation in terms of rôles reveals the invariant factors which are otherwise concealed beneath the changing notation.

The two definitions of number, as a class of classes and as a classification index of signs, seem at first sight hardly comparable, but in fact the difference is not as fundamental as it appears to be. When the Frege-Russell definition is considered in its logical setting, the notion on which it rests, that of class, is found also to be a classification index of signs. The difference between the two definitions is not one of *essence* but of *level*, the level which the definition assigns number in a hierarchy of concepts.

CHAPTER I

THE SENTENCE CALCULUS

Truth Tables. Arithmetical Decision Procedure. Three-valued Logic. Axiomatic Theory. Completeness and Freedom from Contradiction. Independence of the Axioms. Intuitionistic Logic. Polish Bracket-free Notation. Natural Inference.

The sentence calculus

The first work to present mathematics as a deductive system in which the totality of mathematical knowledge is deduced in a purely logical manner from an *explicit* body of assumptions was Euclid's *Elements*. Although by present day standards the *Elements* is entirely wanting in rigour, its definitions pitiful, and its presuppositions (axioms and postulates) woefully inadequate, nevertheless it is of outstanding importance as a *signpost* in the history of the evolution of human thought, pointing the way to a fuller understanding, not only of mathematics, but of reason itself.

One of the respects in which the *Elements* falls short of the requirement of a modern presentation of mathematics is in Euclid's failure to give an account of the logical machinery by which his deductive system operated. We are told about the mathematical presuppositions but not about the logical ones, and are left to extract Euclid's notion of a valid argument from the examples which he gives us. Just as the various intuitively acceptable 'facts' of mathematics are woven together in the *Elements* to produce a coherent structure in which the relations of the parts are made apparent, so it is possible to reveal and codify the 'laws' of reason on which the whole structure rests.

This codification was initiated in modern times by the work of George Boole, who wrote on the mathematical analysis of logic in 1847, though in fact a great deal had been achieved (along modern lines) by the Greek stoics and by the logicians of the Middle Ages.

The foundation of mathematical logic, that is, logic presented in the axiomatic form of Euclid's *Elements*, and, more especially, utilizing the mathematical weapon of variables to

express generality, is the *sentence calculus*. The sentence calculus analyses the structure of compound sentences formed with the connecting terms *and*, *or*, and *implies*, and the prefix *not*. Starting with so-called atomic sentences, concerning which we suppose we have no information save only that we may attribute truth or falsehood to them at will, new sentences are formed by combining atomic sentences with connecting terms, and these compound sentences are tested for truth or falsehood on the various possibilities of truth and falsehood of the constituent atomic sentences.

Denoting atomic sentences by p, q, r, . . ., the connectives *and*, *or*, *implies*, by $\&$, $\vee$, $\rightarrow$ (where $\vee$ is the initial letter of the Latin disjunctive *vel*), and negation by the prefix $\neg$, we consider such compound sentences as $\neg (p \rightarrow q)$, which says that p does not imply q (i.e. q may be false without p being false), and $\{(p \rightarrow q) \,\&\, \neg q\} \rightarrow (\neg p)$ the meaning of which is that if p implies q, and q is not true then p is not true.

Truth tables

With each of the operations $\&$, $\vee$, $\neg$, $\rightarrow$, we associate a truth table according to the intended interpretation. We write T and F as abbreviations for the words *true* and *false* respectively. For negation the table is simply

$\neg$	T | F
	F | T

This table shows that the negation of a true sentence is a false one, and conversely.

The tables for $\&$, $\vee$, $\rightarrow$ are:

$\&$	T	F
T	T	F
F	F	F

$\vee$	T	F
T	T	T
F	T	F

$\rightarrow$	T	F
T	T	F
F	T	T

In the third table, the truth values of the antecedent are in the first column and of the consequent in the first row; the first two tables are symmetrical in rows and columns so that no

distinction is needed. The tables express the fact that a conjunction is true only if both its parts are true, and a disjunction is true if either of its parts are true, and an implication is false only if its antecedent is true and its consequent false. The table for implication is perhaps rather surprising since it marks as true any consequence of a false statement; it is so designed to make $p \to q$ the equivalent of $\sim p \vee q$ which is the sense in which it is generally used in mathematics. The disjunction table is that for *inclusive* disjunction which is true if either or both its alternatives are true.

By means of the truth tables we may test the truth or falsehood of any compound sentence. To illustrate the method we consider a proof of the sentence $((p \,\&\, \sim q) \to q) \to (p \to q)$.

	p	q	$\sim q$	$p \,\&\, \sim q$	$(p \,\&\, \sim q) \to q$	$p \to q$	$((p \,\&\, \sim q) \to q) \to (p \to q)$
1	T	T	F	F	T	T	T
2	T	F	T	T	F	F	T
3	F	T	F	F	T	T	T
4	F	F	T	F	T	T	T

The table shows that the statement is true whatever truth values the atomic sentences p and q may have. The table is constructed as follows; in the first two columns we enter all the four possible choices of truth and falsehood for p and q. Column 3 is formed from column 2 by interchanging T and F according to the table for negation. Column 4 is obtained from columns 1 and 3 by means of the table for conjunction so that we enter a T only in row 2 where there is a T in both columns 1 and 3. Column 5 is obtained from columns 4 and 2 by means of the table for implication, so that an F is entered only in row 2 where there is a T in column 4 and an F in column 2, and columns 6 and 7 are obtained in the same way from columns 1 and 2 and from columns 5 and 6 respectively.

A sentence which is true for all truth values of its atomic parts is said to be universally valid or tautologous; one (like $p \,\&\, \sim p$) which is false for all truth values of its parts is called a contradiction. It is readily verified that the sentence $p \vee \sim p$ is a tautology; it is known as the *tertium non datur* since it affirms

that every sentence must be true or false, and so there is no third possibility (in this logic).

Instead of using truth tables the validity of a compound sentence may be tested by an arithmetical decision procedure. In one such procedure the sentences are variables which take only the values 0, 1 (representing truth and falsehood respectively) and negation, disjunction, conjunction, and implication are represented by the *functions* $1 - p$, pq, $p + q - pq$, and $(1 - p)q$ respectively. These functions take only the values 0, 1 for these values of p and q. The function $1 - p$ takes the *opposite* value to p, pq takes the value 0 if *either* p *or* q takes this value, and takes the value 1 otherwise, $p + q - pq$ takes the value 0 only if p and q are *both* zero, and $(1 - p)q$ takes the value zero if either $p = 1$ or $q = 0$ (and the value 1 only if $p = 0$ and $q = 1$); interpreting 0 and 1 as truth and falsehood respectively the functions associated with the various connectives are thus seen to have the appropriate truth values. The representing function of a tautology is identically zero; for instance the representing function of the tautology $p \vee \sim p$ is $(1 - p)p$ which takes the value zero whether p has the value 0 or the value 1, and representing function of the tautology

$$(p \to q) \to (p \vee r \to q \vee r)$$

is

$$\{1 - (1 - p)q\}(1 - pr)qr$$

which also takes only the value zero whatever values p, q, and r may have. To prove the truth of a complicated sentence we may often shorten the labour of showing that the representing function is identically zero by considering the consequences of supposing that the function takes the value unity.

For instance, to prove that

$$((p \to q) \to r) \to \{(r \to p) \to (s \to p)\}$$

it suffices to observe that if

$$\{1 - (1 - (1 - p)q)r\}\{1 - (1 - r)p\}(1 - s)p = 1$$

then each factor has the value unity, so that $p = 1$, $s = 0$, $r = 1$, and $1 - r = 1$, which is impossible.

Another problem which may very simply be solved by means of representing functions is that of deriving a conclusion from certain assumptions.

For instance, from the assumptions $P \to Q$, $Q \to R$ to derive the conclusion $P \to R$. The assumptions are represented by the equations

$$\text{(i)}\ (1 - P)Q = 0, \quad \text{(ii)}\ (1 - Q)R = 0,$$

where we have written capital instead of small letters, because P, Q, R are not sentence variables which may take either of the values 0, 1 but stand for certain definite sentences (definite numbers, in the representation) which satisfy the assumptions. Multiplying equation (i) by R and (ii) by $1 - P$, and adding, we obtain

$$(1 - P)R((1 - Q) + Q) = 0, \text{ i.e. } (1 - P)R = 0,$$

which represents the desired conclusion $P \to R$. As another example we consider drawing the conclusion T from the assumptions

$$\to P,\ Q,\ (Q \vee R) \to (S \vee T),\ (R \vee S) \to P.$$

The representing equations for the assumptions are

$$1 - P = 0,\ Q = 0,\ (1 - QR)ST = 0,\ (1 - RS)P = 0;$$

substituting these values of P and Q in the third and fourth equations we obtain $ST = 0$ and $1 - RS = 0$, and from the second of these, $R = S = 1$, whence $T = 0$ as required.

We shall see later that to derive a statement P from an hypothesis H is equivalent to proving the statement $H \to P$ without making any assumption, and similarly to derive P from H and K is equivalent to proving $H \to (K \to P)$. Anticipating this result we consider a proof of the statement

$$\{(p \to q) \to r\} \to \{(r \to p) \to (s \to p)\}$$

by drawing the conclusion $s \to p$ from the assumptions

$$(p \to q) \to r,\ r \to p;$$

the assumption equations are $(1 - (1 - p)q)r = 0$, $(1 - r)p = 0$, and adding p times the first to $1 - (1 - p)q$ times the second we find $(1 - (1 - p)q)p = 0$ and so $p = 0$, since $(1 - p)p = 0$, whence finally $(1 - s)p = 0$ which represents the desired conclusion.

If we exhibit the several truth tables in a single table as follows

p	q	$\rightarrow p$	$\rightarrow q$	$p \,\&\, q$	$p \vee q$	$p \rightarrow q$
T	T	F	F	T	T	T
T	F	F	T	F	T	F
F	T	T	F	F	T	T
F	F	T	T	F	F	T

we see that the seven columns of the table introduce seven out of the sixteen possible columns of four T's and F's; the remaining nine arrangements can all be obtained by combinations of the seven sentences p, q, $\rightarrow p$, $\rightarrow q$, $p \,\&\, q$, $p \vee q$, $p \rightarrow q$. The arrangements TTTT and FFFF are obtained for instance with the sentences $p \vee \rightarrow p$ and $p \,\&\, \rightarrow p$. Of the four arrangements with a single T we have already obtained one by $p \,\&\, q$ and the remainder correspond to $p \,\&\, \rightarrow q$, $\rightarrow p \,\&\, q$, and $\rightarrow p \,\&\, \rightarrow q$; the negations of these sentences generate the four arrangements with a single F. Of the six arrangements with two T's, four are exhibited in the first four columns above and the remaining two are generated by the sentence

$$(p \,\&\, q) \vee (\rightarrow p \,\&\, \rightarrow q)$$

and its negation. *It follows that there can be no logical connective independent of* &, $\vee$, $\rightarrow$ *and* $\rightarrow$. Moreover, not even these are independent, since $p \rightarrow q$ has the same truth values as $\rightarrow p \vee q$ and $p \,\&\, q$ the same truth values as $\rightarrow (\rightarrow p \vee \rightarrow q)$; in fact just a single connective suffices, for if $p|q$ denotes *not both p and q* then $\rightarrow p$ has the same truth values as $p|p$, $p \,\&\, q$ the same as $(p|q)|(p|q)$, $p \vee q$ the same as $(p|p)|(q|q)$, and $p \rightarrow q$ the same as $\{(p|p)|(p|p)\}|(q|q)$.

The truth tables for the logical operations may be looked at from two points of view; we may regard the operators &, $\vee$, $\rightarrow$, $\rightarrow$, as *abbreviations* for the signs of known concepts and construct the truth tables with this interpretation before us, as in fact we did above, but we may also take the truth tables as *the definitions of the operations* &, $\vee$, $\rightarrow$, $\rightarrow$, without presupposing the interpretation of the signs. It is this second viewpoint which is important in the sequel.

Three-valued logic

The construction of logic by means of postulated truth tables sets logic free from the chains in which it is bound by habits of thought and opens the door to generalized logics with more than two truth values. Such generalized logics have not yet found any important applications but have a structure no less interesting than that of the familiar two-valued logic. We content ourselves with a brief mention of a logic with three truth values, *true*, *indeterminate*, and *false*, denoted by T, I, and F respectively, and the operations $\neg$, &, $\vee$, $\rightarrow$, defined by the following tables

$\neg$	
T	F
I	I
F	T

&	T	I	F
T	T	I	F
I	I	I	F
F	F	F	F

$\vee$	T	I	F
T	T	T	T
I	T	I	I
F	T	I	F

$\rightarrow$	T	I	F
T	T	I	F
I	I	I	I
F	T	T	T

If we assign to T, I, F, the numerical values 2, 1, 0, and associate with each operation a numerical function, $\neg(x)$, $\&(x,y)$, $\vee(x,y)$, and $\rightarrow(x,y)$ denoting the functions associated with $\neg$, &, $\vee$, and $\rightarrow$ respectively, (where $\neg(x)$ is the truth value of the negation of a sentence with truth value x, $\rightarrow(x,y)$ the truth value of an implication whose antecent has truth value x and consequent truth value y, etc.) then the information contained in the tables may be summed up in the definitions,

$$\neg(x) = x + 2 \pmod 3, \quad \&(x,y) = \min(x,y),$$
$$\vee(x,y) = \max(x,y), \quad \rightarrow(x,y) = y[x/2] + 2 - x,$$

(where min (x,y), max (x,y) denote the lesser and greater of x,y respectively and $[x/2]$ denotes the whole part of half x). By contrast with two-valued logic, in this three-valued logic all the operations &, $\vee$, $\multimap$, $\rightarrow$, are independent, $\multimap \multimap p$ is different from p, and the *tertium-non-datur* $p \vee \multimap p$ does not hold; instead we have the *quartum-non-datur* $p \vee \multimap p \vee \multimap \multimap p$.

Axiomatic theory

In view of the great importance of the study of axiomatic systems we shall now contrast the formulation of the sentence calculus by means of truth tables with its formulation as an axiomatic system.

The axiomatic theory is obtained by selecting certain (universally valid) initial sentences as axioms and providing rules of procedure whereby all the remaining universally valid sentences may be deduced. (Of course the notion of universal validity is now an extraneous one and finds no expression inside the axiomatic theory.) We use the letters p, q, r, . . . now, not as particular atomic sentences, but as variables for which sentences may be substituted. As sentences we take the variables themselves, and combinations of sentences with the operators $\multimap$ and $\vee$, like $\multimap p$, $p \vee q$, and $\multimap (p \vee q) \vee q$. We shall continue to use the signs & and $\rightarrow$, but only as abbreviations, not as elements of the axiomatic systems. Thus $p \,\&\, q$ is an *abbreviation* for $\multimap (\multimap p \vee \multimap q)$ and $p \rightarrow q$ is an abbreviation for $\multimap p \vee q$. To express the fact that by the rules of procedure (to be specified) we can pass from one or more sentences $p, q, \ldots$ to a sentence r we shall write $p, q, \ldots \vdash r$ (*read* r follows from $p, q, \ldots$); if p is an axiom, or follows from an axiom we write $\vdash p$ (where the sign $\vdash$ now has no antecedent). We shall consider the system based on the four axioms

$$\text{(i)} \vdash p \vee p \rightarrow p, \quad \text{(ii)} \vdash p \rightarrow p \vee q, \quad \text{(iii)} \vdash p \vee q \rightarrow q \vee p,$$
$$\text{(iv)} \vdash (p \rightarrow q) \rightarrow (r \vee p \rightarrow r \vee q)$$

(to save parentheses, adopting the convention that $\vee$ has priority over $\rightarrow$, i.e. that $r \vee p \rightarrow r \vee q$ stands for

$$(r \vee p) \rightarrow (r \vee q), \text{ not } r \vee (p \rightarrow r) \vee q)$$

The derivation of q from p will become a proof of $p \to q$ by this transformation provided that

(i) if A is universally valid then so is $p \to A$,

(ii) the scheme

$$\frac{p \to A \qquad p \to (A \to B)}{p \to B}$$

is valid.

As to (i) the sentence $p \to A$ may be derived, by *modus ponens*, from A and the universally valid sentence $A \to (p \to A)$; the validity of (ii) is established by means of the universally valid sentence

$$(p \to A) \to [\{p \to (A \to B)\} \to (p \to B)]$$

which by *modus ponens* takes us from $p \to A$ to

$$\{p \to (A \to B)\} \to (p \to B)$$

by which, by another application of *modus ponens* we derive $p \to B$ from $p \to (A \to B)$.

It readily follows that if we can prove r by adding p and q to the axioms then we can derive $p \to (q \to r)$ from axioms (i) to (iv) alone, for we may first transform the derivation of r from p and q into a derivation of $q \to r$ from $q \to p$ and in turn this derivation may be transformed into a proof of $p \to (q \to r)$, (since $p \to (q \to p)$ is universally valid), or we may directly transform the derivation of r into a proof of $(p \& q) \to r$ which is equivalent to $p \to (q \to r)$. The result extends of course to any number of statements $p_1, p_2, \ldots, p_n$ in the form: if q is derivable from axioms (i) to (iv) and $p_1, p_2, \ldots, p_{n-1}, p_n$ then $p_n \to q$ is derivable from the axioms and $p_1, p_2, \ldots, p_{n-1}$ and $p_1 \to (p_2 \to \ldots (p_n \to q) \ldots)$ is derivable from axioms (i) to (iv).

Completeness and freedom from contradiction

It can be readily shown that every sentence provable in the axiom system is universally valid and conversely that every universally valid sentence (of the calculus with two truth values) is provable in the axiomatic system. For this reason the axiom system is said to be *complete with respect to the truth tables*; it is *complete* also in a stricter sense, since it can be shown that if any

sentence A, which is not provable from the axioms (i) to (iv), is added to the axioms, the resulting system contains a contradiction. For instance if we add the sentence $p \vee q$ to the axioms then we can prove $p \vee p$, and $\rightarrow p \vee \rightarrow p$ and from these we may derive both p and $\rightarrow p$ which is a contradiction. That no contradiction can be derived from the axioms (i) to (iv) themselves may be shown as follows. Testing the axioms (i) to (iv) by the truth tables we readily verify that each of them is universally valid. Moreover if p and $p \rightarrow q$ are universally valid then so is q (for the truth table shows that $T \rightarrow F$ is false) and if we replace a variable p in a universally valid sentence, say $S(p)$ by another sentence, A (say), then the resulting sentence $S(A)$ is also universally valid since $S(p)$ is true both when p is true and when p is false. It follows that any sentence derived by the derivation rule from the axioms is universally valid. Accordingly, if p is derivable then p is universally valid so that $\rightarrow p$ is not universally valid and therefore $\rightarrow p$ is not derivable from the axioms.

Independence of axioms

Another important property of the axioms (i) to (iv) is their *independence*; none of the four axioms is derivable from the remainder. The proof of independence is similar to the foregoing proof of consistency. For example to prove the independence of axiom (ii) we consider an interpretation in which sentences have four values 0, 1, 2, 3, and the operators $\rightarrow$, $\vee$, the following tables:

$\rightarrow$	0	1	2	3
	1	0	3	2

$\vee$	0	1	2	3
0	0	0	0	0
1	0	1	1	1
2	0	1	2	2
3	0	1	2	3

For all values of the variables, axioms (i), (iii), (iv) assume only the values 0 and 2, and since the property of taking only these values is preserved under substitution and under *modus ponens* (since $1 \vee q = 0$ only if $q = 0$, $3 \vee q = 0$ only if $q = 0$ and $3 \vee q = 2$ only if $q = 2$) it follows that all sentences derived from axioms (i), (iii), (iv) take only these values. But axiom (ii) has the value 1 for the values $p = 2$, $q = 1$.

Intuitionistic logic

We saw that one of the features of a three-valued logic is that the *tertium-non-datur* does not hold in it. There is, however, a two-valued system of logic in which the *tertium-non-datur* does not hold although many other familiar rules of reasoning remain valid. This system is known as *intuitionistic logic* (from its connection with a system of thought which finds no other source for mathematical concepts and proof methods than intuition) and is important in connection with some paradoxes to which we shall refer later. By contrast with intuitionistic logic, the two-valued system based on axioms (i) to (iv) is called *classical logic*. In intuitionistic logic the operators &, $\vee$, $\rightarrow$, $\leftrightarrow$, are all independent. A set of axioms for intuitionistic logic is:

1. $p \rightarrow (q \rightarrow p)$
2. $\{p \rightarrow (q \rightarrow r)\} \rightarrow \{(p \rightarrow q) \rightarrow (p \rightarrow r)\}$
3. $p \rightarrow (q \rightarrow (p \,\&\, q))$
4. $(p \,\&\, q) \rightarrow p$
5. $(p \,\&\, q) \rightarrow q$
6. $p \rightarrow (p \vee q)$
7. $q \rightarrow (p \vee q)$
8. $(p \rightarrow r) \rightarrow \{(q \rightarrow r) \rightarrow ((p \vee q) \rightarrow r)\}$
9. $(p \rightarrow q) \rightarrow \{(p \rightarrow \neg q) \rightarrow (\neg p)\}$
10. $(\neg p) \rightarrow (p \rightarrow q)$

The derivation rules are the same as for classical logic. Each of these axioms is universally valid under the two-valued truth table interpretation, and such sentences as

$$p \rightarrow (\neg (\neg p)),\ \neg (\neg (\neg p)) \rightarrow (\neg p),\ \neg (\neg p) \vee \neg p$$

are derivable, but $p \vee \neg p$ is not a consequence of the intuitionistic axioms, nor is $(\neg \neg p) \to p$. If $p \vee \neg p$ is postulated as an eleventh axiom the enlarged system is fully equivalent to classical logic and every universally valid sentence is then derivable.

Classical logic and intuitionistic logic are very intimately related. Every consequence of the intuitionistic axioms is derivable from the classical axioms (being universally valid); on the converse side it is known that if $\neg$ A is provable in classical logic (for some sentence A) then $\neg$ A is provable in intuitionistic logic, and further that if E is a sentence formed from the operators & and $\neg$ only and if E is provable classically then it is provable intuitionistically; in fact if we transform any sentence E into a classically equivalent sentence E′ by the transformations

$$\begin{array}{cccc} A \to B & A \,\&\, B & A \vee B & \neg A \\ \neg (A \,\&\, \neg B) & A \,\&\, B & \neg (\neg A \,\&\, \neg B) & \neg A \end{array}$$

then when E is provable classically, E′ is provable intuitionistically.

The class of provable sentences in classical logic (axioms (i) to (iv)) is the same as the class of universally valid sentences of the two-valued truth tables, but no truth table with a finite number of values has for its class of universally valid sentences, the provable sentences of intuitionistic logic. In other words intuitionistic logic does not admit an interpretation in terms of truth tables with a finite number of truth values. It is, however, possible to give a truth table interpretation with an infinite number of truth values.

Bracket free notation

Logical notations have been devised which entirely obviate the need for brackets. The two-valued classical logic, for instance, with the operations C (for implication) and N (for negation) may be formulated with the axioms

(*a*) C C p q C C q r C p r, (*b*) C C N p p p, (*c*) C p C N p q

and the following rules of operation:

(i) For any small letter (in an axiom or proved sentence) at all its points of occurrence we may write either another

small letter, or a pair in which the first member is N and the second any small letter, or a trio in which the first is C and the other two are small letters.

(ii) From the sentences A, C A B we may derive the sentence B.

The only sentences are those which may be formed from the axioms by the rules of substitution (i). To 'read' a sentence in this notation we look to the last capital letter first and pick out the constituent sentences by passing from capital to capital in reverse order. For instance in axiom (*a*) we first isolate the final C with its arguments *p*, *r*, then the previous C with its arguments *q*, *r*, forming the arguments

$$(\mathrm{C}\,q\,r) \text{ and } (\mathrm{C}\,p\,r)$$

of the C which precedes them; this C with its arguments forms the argument C (C *q r*) (C *p r*) of the first C, the other argument being C *p q*. Writing $p \to q$ for C *p q* this analysis shows that axiom (*a*) is equivalent to

$$(p \to q) \to \{(q \to r) \to (p \to r)\}$$

Similarly, writing $\sim p$ for N *p*, axiom (*b*) is equivalent to

$$(\sim p \to p) \to p$$

which we obtain by looking first at the final capital N which attaches to the following *p* forming one argument N *p* of the preceding C; in this way we isolate the part C (N *p*) *p*, i.e. $\sim p \to p$ which forms the first argument of the first C. In the same way we can translate axiom (*c*) into $p \to (\sim p \to q)$. To illustrate the use of the notation we outline the proof of the sentence C *p p*.

Substituting *p* for *q* in axiom (*c*) we obtain the sentence

$$\mathrm{C}\,p\,\mathrm{C}\,\mathrm{N}\,p\,p,$$

and substituting *p* for *r* and C N *p p* for *q* in axiom (*a*) we obtain

$$\mathrm{C}\,\mathrm{C}\,p\,\mathrm{C}\,\mathrm{N}\,p\,p\,\mathrm{C}\,\mathrm{C}\,\mathrm{C}\,\mathrm{N}\,p\,p\,p\,\mathrm{C}\,p\,p;$$

this sentence is formed by an initial C followed by the proved sentence C *p* C N *p p* followed by the sentence C C C N *p p p* C *p p* and so C C C N *p p p* C *p p* is proved. This in turn is

formed from an initial C, the axiom C C N p p p and the sentence C p p, which shows that this last is a proved sentence.

Natural inference

All the systems of logic which we have so far described bear little resemblance to the familiar processes of mathematical argument. They do not allow, for instance, the introduction of hypotheses (except in an application of the deduction theorem) and are both slow in action and inflexible. About twenty years ago a far more powerful formulation of logic was discovered. This formulation is known as *natural inference*; in natural inference there are no axioms, only a series of derivation rules. These rules are separated into *introduction* and *elimination* rules.

The introduction rules are:

$$\frac{A, B}{A \,\&\, B}\ \& \qquad \frac{A}{A \vee B}\ \frac{B}{A \vee B}\ \vee$$

$$\frac{A \vdash B}{A \to B}\ \to \qquad \frac{A \vdash B, \neg B}{\neg A}\ \neg$$

The elimination rules are:

$$\frac{A \,\&\, B}{A}, \frac{A \,\&\, B}{B}\ \& \qquad \frac{A \vee B, A \vdash C, B \vdash C}{C}\ \vee$$

$$\frac{A, A \to B}{B}\ \to \qquad \frac{\neg \neg A}{A}\ \neg .$$

We have written the operator involved in each inference beside the inference rule to serve as a *name* for the rule. Thus the rule $\frac{A, B}{A \,\&\, B}$ which says that from A and B we may infer A & B is called *and- introduction* and the rule $\frac{A \vee B, A \vdash C, B \vdash C}{C}$ which says that from the assumptions A $\vee$ B, and proofs that C follows from A, and C follows from B, we may infer C, is called *or-elimination*. The rule for *not-introduction* says that if we have a derivation of B from A then we may infer $\neg$ A from $\neg$ B. Unlike a rule of inference in an axiomatic system, which is applied only to derive a proved sentence from a proved

sentence, in natural inference, a rule of inference may be applied to *any* sentence. A sentence which follows from a *proved* sentence by a rule of inference, is thereby proved, but a sentence inferred from an unproved sentence is itself unproved. Since there are no axioms in natural inference, proof must always be based on the rule for *implication-introduction* which says that A → B is *proved* if B can be inferred from A by the rules of inference. In a proof in an axiomatic system every sentence is either an axiom or an immediate consequence of one of the preceding sentences in the proof by a rule in inference. In a proof by natural inference we start with hypotheses (i.e. unproved sentences), explore their consequences by the rules of inference and then utilize information of the sort A ⊢ B (i.e. B follows from A) to infer that A → B is *proved*.

Simply by omitting the rule $\dfrac{\neg\neg A}{A}$ for *not-elimination* we obtain a natural inference formulation of intuitionistic logic. (The use of the derivation symbol A ⊢ B allows either A or B to be absent; ⊢ B says that B is derivable without assumption and so *provable*, and A ⊢ denotes that any sentence is derivable from A, and so A is false.) Of course the inference rules (when applied to proved sentences) are all provable from the appropriate axiom system (or, like *modus ponens*, is itself an inference rule of the axiom system). The rule for *implication-introduction*, for instance, is proved by the deduction theorem, and the rule for *or-elimination* is a consequence of the universally valid sentence

$$\{(A \vee B) \mathbin{\&} (A \rightarrow C) \mathbin{\&} (B \rightarrow C)\} \rightarrow C.$$

As an illustration of the technique we prove the universally valid sentence

$$\{A \vee (B \mathbin{\&} C)\} \rightarrow \{(A \vee B) \mathbin{\&} (A \vee C)\}$$

by the method of natural inference.

$$\dfrac{\dfrac{A \vee (B \mathbin{\&} C) \quad \dfrac{\dfrac{A}{A \vee B}\vee \quad \dfrac{A}{A \vee C}\vee}{(A \vee B) \mathbin{\&} (A \vee C)}\& \quad \dfrac{\dfrac{\dfrac{B \mathbin{\&} C}{B}\&}{A \vee B}\vee \quad \dfrac{\dfrac{B \mathbin{\&} C}{C}\&}{A \vee C}\vee}{(A \vee B) \mathbin{\&} (A \vee C)}\&}{(A \vee B) \mathbin{\&} (A \vee C)}\vee}{\{A \vee (B \mathbin{\&} C)\} \rightarrow \{(A \vee B) \mathbin{\&} (A \vee C)\}}\rightarrow$$

Explanation

Starting at the top right hand side we proceed to derive consequences from the hypothesis B & C. By and-elimination we derive first the consequences B and C, and then by or-introduction we derive both A ∨ B and A ∨ C and finally, by and-introduction, (A ∨ B) & (A ∨ C). This constitutes a derivation of (A ∨ B) & (A ∨ C) from (B & C). In the centre column we find a derivation of (A ∨ B) & (A ∨ C) from A by or-introduction and and-introduction. Thus on the long horizontal line we have first the assumption A ∨ (B & C), then a derivation of (A ∨ B) & (A ∨ C) from A and a derivation of the same sentence from B & C. By or-elimination we obtain a derivation of (A ∨ B) & (A ∨ C) from A ∨ (B & C) and thence, by implication-introduction, we *prove* the desired sentence

$$\{A \vee (B \,\&\, C)\} \rightarrow \{(A \vee B) \,\&\, (A \vee C)\}$$

(It should be noted that the names attached to the inference steps are for reference only and are not part of the proof by natural inferences.) A proof of the same sentence in classical logic (axioms (i) to (iv)) fills several pages but the shortest proof is by means of representing functions.

CHAPTER II

PREDICATE CALCULUS

Axiomatic Theory. Freedom from Contradiction. The Deduction Theorem. Natural Inference. Validity and Satisfiability. Decision Procedure for Monadic Predicates. Gödel's Completeness Theorem.

Predicate calculus

The sentence calculus, in any of the forms we have described, is the basis of all mathematical logic but it is inadequate for arithmetic. In the sentence calculus we operated upon sentences as unanalysed wholes; we turn now to the logical analysis of sentences into subject and predicate as the second step in the construction of a logic adequate for arithmetic. We denote subjects of sentences by small letters, called individual variables, and predicates by capitals attached to small letters, in the manner of function signs in mathematics, such as $A(x)$, $B(x,y)$. Universality (over individuals, not predicates) is expressed by the operator $\forall$, and existence by the operator $\exists$. To formulate *predicate logic* we must add to the axioms (and inference rules) of sentence logic the two *additional* rules of inference

$$\mathrm{PI}_1 \quad \frac{\mathrm{S} \to \mathrm{A}(t)}{\mathrm{S} \to \forall t \mathrm{A}(t)}, \qquad \mathrm{PI}_2 \quad \frac{\mathrm{A}(t) \to \mathrm{S}}{\exists t \mathrm{A}(t) \to \mathrm{S}},$$

and the axioms,

$$\mathrm{PA}_1 \quad \forall x \mathrm{A}(x) \to \mathrm{A}(t), \qquad \mathrm{PA}_2 \quad \mathrm{A}(t) \to \exists x \mathrm{A}(x).$$

In predicate logic we have sentences like $\forall x \mathrm{A}(x)$ in which the variable x is bound by the operator $\forall$, and also free variable sentences like $\mathrm{A}(x)$ in which the variable x does not come within the scope of one of the operators $\forall$ and $\exists$. Sentences are combined to form new sentences exactly as in sentence logic. In the new, additional, rules of inference S stands for *any* sentence in which t does not appear as a free variable, and $\mathrm{A}(t)$ is *any* sentence in which t appears free. The notation is not intended to preclude the possibility of applying the rule of inference when S and $\mathrm{A}(t)$ contain other variables (not written

down) different from t. In the two axioms, t is any variable, and $A(x)$ may contain variables other than x (even t).

To sum up, we have to distinguish in predicate logic three classes of variables:

(i) Sentence variables $p, q, r, \ldots$
(ii) Individual variables $x, y, z, \ldots$
(iii) Predicate variables $A(x)$, $A(x,y)$, $B(x)$, . . .

A *sentence* of predicate logic is either a sentence variable, a predicate variable, or a combination of sentences $\neg S$, $S \,\&\, T$, $S \vee T$, $S \rightarrow T$, (where the sentences S, T do not contain a variable free in one and bound in the other), or is one of the forms $\forall x S(x)$, $\exists x S(x)$ where $S(x)$ is a sentence in which x is a free variable.

Substitution in variables is more complicated in predicate logic than in sentence logic owing to the duality of free and bound variables. Any predicate may be substituted for a sentence variable, a free individual variable (at all its points of occurrence) may be replaced by another which does not already occur bound in the sentence, and a bound individual variable may be replaced by another (at all its points of occurrence) provided that the change is simultaneously made in the operator involved, and that the new variable does not already occur free in the sentence.

As examples of the kind of sentences provable in predicate logic we may mention

$$(\forall x)\ (A(x) \rightarrow B(x)) \rightarrow \{(\forall x)A(x) \rightarrow (\forall x)B(x)\}$$
$$(\forall x)\ (A(x) \rightarrow B(x)) \rightarrow \{(\exists x)A(x) \rightarrow (\exists x)B(x)\}$$

and
$$(\exists x)\ (\forall y)A(x,y) \rightarrow (\forall y)\ (\exists x)A(x,y);$$

to illustrate the proof technique of predicate logic we give the details of the proof of the first of these sentences. The starting point of the proof is the group of universally valid sentences

(*a*) $\{p \rightarrow (q \rightarrow r)\} \rightarrow \{q \rightarrow (p \rightarrow r)\}$
(*b*) $\{p \rightarrow (q \rightarrow r)\} \rightarrow \{(p \,\&\, q) \rightarrow r\}$
(*c*) $\{(p \,\&\, q) \rightarrow r\} \rightarrow \{p \rightarrow (q \rightarrow r)\}$.

By means of (*b*) and (*c*) we may generalize the rule of inference PI_1 to provide the rule

$$\frac{S \rightarrow (T \rightarrow A(x))}{S \rightarrow (T \rightarrow \forall x A(x))};$$

for by (*b*) and *modus ponens* we derive $(S \& T) \to A(x)$ from $S \to (T \to A(x))$ then by PI_1 we obtain $S \& T \to (\forall x)A(x)$, and finally from (*c*) and *modus ponens* we derive $S \to (T \to (\forall x)A(x))$. (A similar argument shows that the rule of inference is valid also with more than two implications). By axiom PA_1 we have

$$(\forall x)(A(x) \to B(x)) \to (A(t) \to B(t)),$$

whence by substituting in (*a*), and using *modus ponens*,

$$A(t) \to \{(\forall x)(A(x) \to B(x)) \to B(t)\};$$

then by axiom PA_1 again, and *modus ponens*,

$$(\forall x)A(x) \to \{(\forall x)(A(x) \to B(x)) \to B(t)\}$$

whence by the generalized rule of inference we reach

$$(\forall x)A(x) \to \{(\forall x)(A(x) \to B(x)) \to (\forall x)B(x)\}$$

and by a further use of (*a*), and *modus ponens*, we complete the proof of the sentence

$$(\forall x)\{A(x) \to B(x)\} \to \{(\forall x)A(x) \to (\forall x)B(x)\}.$$

Like sentence logic, predicate logic is demonstrably consistent and its axioms independent. To prove consistency we consider an interpretation of predicate logic in which both sentence variables and predicate variables take just two values T and F (irrespective of how the argument places in the predicate signs are filled). Both $(\forall x)A(x)$ and $(\exists x)A(x)$ are identified with A. This interpretation is equivalent to the assumption that there is only one individual, *i* say, so that $(\forall x)A(x)$ and $(\exists x)A(x)$ both affirm $A(i)$, which may be either true or false. The tables for $\to$ and $\vee$ are the usual ones of two-valued logic. Under this interpretation it can readily be seen that all the axioms have the value T and that the rules of inference yield only sentences of value T from sentences of value T. Hence if a sentence A is provable then its value is T, and so the value of $\neg A$ is F and therefore $\neg A$ is *not* provable.

The sentence calculus, we saw, is complete in two different senses of the word. On the one hand it is complete because the addition to the axioms of any unproved formula makes the system contradictory; this kind of completeness is not shared by the predicate calculus, since there are unprovable sentences which are not inconsistent with the axioms. For instance the

sentence $(\exists x)A(x) \to (\forall x)A(x)$ is true if there is only a single individual, so that the sentence is not inconsistent with the axioms, but it is not a consequence of the axioms as may be shown by considering an interpretation in which there are two individuals i, j, and $(\exists x)A(x)$, $(\forall x)A(x)$, are taken to be $A(i) \vee A(j)$ and $A(i)$ & $A(j)$ respectively.

The second sense in which the sentence calculus is complete is that every universally valid sentence is provable, and in this sense the predicate calculus is also complete. However the task of specifying the universally valid sentences of the predicate calculus is one which we shall postpone until we have briefly considered an extension of the method of natural inference to the predicate calculus.

The deduction theorem is valid for the predicate calculus in the form:

If B is derivable by the predicate calculus from the axioms of the predicate calculus and the additional axioms $A_1, A_2, \ldots, A_k$ and A and if there is no use made of a free variable in A in this derivation, then $A \to B$ is derivable from $A_1, A_2, \ldots, A_k$.

If the restriction on the use of free variables is satisfied for all of $A_1, A_2, \ldots, A_k$ then it follows that

$$A_1 \to (A_2 \to (A_3 \to (\ldots A \to B \ldots)) \text{ is provable,}$$

i.e. $$(A_1 \,\&\, A_2 \,\&\, \ldots \,\&\, A) \to B \text{ is provable.}$$

To prove the deduction theorem, we shall suppose first that A does not contain free a variable which occurs (as t) in an application of PI_1 or PI_2 in the derivation of B. Replace each sentence S in the derivation of B by $A \to S$; we show that the resulting system of sentences constitutes a derivation of $A \to B$ from $A_1, A_2, \ldots, A_k$. Each axiom Σ (of predicate logic) in the derivation of A becomes $A \to \Sigma$ which is derivable from Σ; so too each hypothesis A_i becomes $A \to A_i$ which is derivable from A_i. An application of *modus ponens*

$$\frac{\begin{array}{c}S\\S \to T\end{array}}{T}$$

becomes

$$\frac{\begin{array}{c}A \to S\\A \to (S \to T)\end{array}}{A \to T}$$

which we have already seen to be a valid schema (in the proof of the deduction theorem for sentence logic).

An application of PI_1 in the form

$$\frac{P \to Q(t)}{P \to \forall t Q(t)}$$

becomes

$$\frac{A \to (P \to Q(t))}{A \to (P \to \forall t Q(t))}$$

which is valid,

for the hypothesis is equivalent to $A \,\&\, P \to Q(t)$, from which we may derive $A \,\&\, P \to \forall t Q(t)$ by PI_1 itself, since A does not contain t as a free variable, and $A \,\&\, P \to \forall t Q(t)$ is equivalent to $A \to (P \to \forall t Q(t))$. Similarly an application of PI_2,

$$\frac{Q(t) \to P}{\exists t Q(t) \to P}$$

becomes

$$\frac{A \to (Q(t) \to P)}{A \to (\exists t Q(t) \to P)}$$

which is also seen to be valid since the hypothesis is equivalent to $Q(t) \to (A \to P)$ from which we may derive $\exists t Q(t) \to (A \to P)$, by an application of PI_2, and this in turn is equivalent to $A \to (\exists t Q(t) \to P)$, as desired.

If A in fact contains a variable which plays the part of t in an application of PI_1 or PI_2 in the derivation of B then we replace this variable at each point of its occurrence in A, but not elsewhere, by a new variable, w say, which does not occur at all in the derivation of B; let the sentence which results from this substitution in A be denoted by A*.

The hypothesis that no free variable in A is *used* in the derivation of B means that the replacement of t in $A(t)$ by w, does *not* invalidate the derivation of B, and so the result of this replacement is to transform a derivation of B from A, A_1, A_2, . . . A_k into a derivation of B from A*, A_1, A_2, . . ., A_k; it follows that $A^* \to B$ is derivable from A_1, A_2, . . ., A_k and hence, substituting t for w, in $A^* \to B$, we achieve a derivation of $A \to B$ from A_1, A_2, . . ., A_k as required.

The validity of the deduction theorem for the predicate calculus makes it possible to extend the method of natural inference to predicate logic. The extension we are going to

consider is in fact independent of the formulation of a method of natural inference for the sentence calculus.

In this formulation a deduction by natural inference is a finite series of sentences each of which is either an hypothesis or may be inferred from an earlier sentence by one of the following rules of inference. In stating the rules of inference we suppose that in any particular deduction the variables have been assigned some particular order (e.g. x, x', x'', . . ., or perhaps some alphabetical variation).

The first rule is that a sentence ψ may be inferred from earlier sentences $\phi_1, \phi_2, \ldots, \phi_k$ if $\Sigma\phi \rightarrow \psi$ is provable in the sentence calculus (by any method we please), where $\Sigma\phi$ stands for $\phi_1 \,\&\, \phi_2 \,\&\, \ldots \,\&\, \phi_k$; this may be called rule T (by reference to *tautology*).

In a deduction by natural inference any hypothesis introduced may be discharged by an application of the schema of implication-introduction (II)

$$\frac{\phi \quad \psi}{\phi \rightarrow \psi};$$

in the application of this schema ϕ is a hypothesis, and each new hypothesis introduced between the line on which ϕ stands and the line on which ψ stands must be discharged before the line on which ψ stands. Hypotheses are numbered and beside the line in which hypothesis number n is discharged we write $-n$. In a correct application of the schema (II) the sum of the numbers of all the lines from ϕ to $\phi \rightarrow \psi$ is zero. In the deduction of $\phi \rightarrow \psi$ from $\phi \vdash \psi$, ψ is called the last line of the deduction and ϕ is called the premiss.

The remaining rules of inference are the rules for $\forall$ and $\exists$ introduction and elimination.

Introduction	Elimination
$\dfrac{\phi(\beta)}{(\forall\alpha)\phi(\alpha)}\ (\forall)$	$\dfrac{(\forall\alpha)\phi(\alpha)}{\phi(\beta)}\ (\forall)$
$\dfrac{\phi(\beta)}{(\exists\alpha)\phi(\alpha)}\ (\exists)$	$\dfrac{(\exists\alpha)\phi(\alpha)}{\phi(\beta)}\ (\exists)$

In each rule of inference $\phi(\beta)$ denotes the result of substituting

β for α at each point of its occurrence in $\phi(\alpha)$. Both in V-introduction and in $\exists$-elimination the schema is applicable only if β denotes a variable which is preceded by all the free variables in $(\forall\alpha)\phi(\alpha)$. An application of V-introduction or of $\exists$-elimination *restricts* the variable β, and a variable once restricted must not be used again in either of these inferences. A deduction by natural inference of ψ from ϕ is not complete if a *restricted* variable is free in either ϕ or ψ.

The operation of the method is easily shown by examples. As a first example we consider a proof of the sentence

$$(\exists x)(\forall y)\mathrm{F}(x,y) \rightarrow (\forall y)(\exists x)\mathrm{F}(x,y).$$

We take the natural order of the variables to be w, x, y, z, t. The proof is as follows.

1. $(\exists x)(\forall y)\mathrm{F}(x,y)$	Hypothesis
$(\forall y)\mathrm{F}(x,y)$, x restricted,	by $\exists$-elimination
$\mathrm{F}(x,y)$	by V-elimination
$(\exists x)\mathrm{F}(x,y)$	by $\exists$-introduction
$(\forall y)(\exists x)\mathrm{F}(x,y)$	by V-introduction
— 1. $(\exists x)(\forall y)\mathrm{F}(x,y) \rightarrow (\forall y)(\exists x)\mathrm{F}(x,y)$	by (II)

The converse of this sentence is false, and it is interesting to see how the restrictions on the schema prevent the deduction. The attempted proof proceeds as follows

1. $(\forall x)(\exists y)\mathrm{F}(x,y)$	Hypothesis
$(\exists y)\mathrm{F}(w,y)$	V-elimination
$\mathrm{F}(w,z)$, z restricted,	$\exists$-elimination

Here the proof breaks down, since the attempt to infer $(\forall x)\mathrm{F}(x,z)$ is disallowed as w is followed by the free variable z in $(\forall x)\mathrm{F}(x,z)$. If, instead of w, we introduce a later letter z in the second line, giving

$$(\exists y)\mathrm{F}(z,y)$$

then we are unable to infer $\mathrm{F}(z,w)$ since w precedes z, but must take instead $\mathrm{F}(z,t)$, and we can make no further progress towards $(\forall x)\mathrm{F}(x,t)$ since z precedes t.

As another example we consider a proof of the sentence

$$(\forall x)(F(x) \to G(x)) \to \{(\exists x)F(x) \to (\exists x)G(x)\}.$$

1.	$(\forall x)(F(x) \to G(x))$	Hypothesis
	$F(x) \to G(x)$	$\forall$-elimination
2.	$(\exists x)F(x)$	Hypothesis
	$F(x)$, x restricted,	$\exists$-elimination
	$G(x)$	from lines 2, 4 by Γ
	$(\exists x)G(x)$	by $\exists$-introduction
— 2.	$(\exists x)F(x) \to (\exists x)G(x)$	by (II)
— 1.	$(\forall x)(F(x) \to G(x)) \to \{(\exists x)F(x)$	
	$\to (\exists x)G(x)\}$	by (II)

It is known that every provable sentence of the predicate calculus is provable by this method of natural inference, and conversely that every sentence provable by natural inference is provable in the predicate calculus.

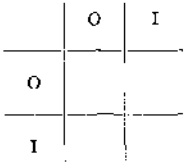
Valid and satisfiable predicates

To provide a content for statements of the predicate calculus we suppose given some domain of individuals to which the individual variables, and the universal and existential operators refer. This domain may be finite or infinite, but not empty. A definite predicate (value of a predicate variable) defined for a given domain is simply an assignment of a truth value for each assignment of individuals from the domain for the individual variables in the predicate. Thus, for instance, in a domain of two individuals 0,1 say, there are just four predicates with one argument, sixteen with two arguments, 256 with three arguments, and so on; each predicate with two arguments is given by a table like

	0	1
0	T	F
1	F	T

The predicate given by this table is true for the (ordered) pair of values (0,0), false for the pair (1,0), and so on. The tables for

the whole group of sixteen predicates with two arguments are obtained by filling in the spaces in the schema



with T's or F's in all possible ways.

A statement in predicate logic is said to be *valid* in a given domain if it takes the value *true* for every substitution of definite predicates for the predicate variables, and individuals (from the given domain) for the individual variables. For instance, $(\exists x)P(x) \rightarrow (\forall x)P(x)$ is valid in a domain of only one individual, but not valid in a domain of two individuals (since the truth of P for one of the individuals does not make every P true also for the second individual). A statement is said to be *satisfiable* in a given domain if we can make a selection of predicates for the predicate variables, and a selection of individuals (from the domain) for the individual variables, for which the statement is true. For example, $(\exists x)(\exists y)\{P(x) \ \& \ \neg P(y)\}$ is satisfiable in a domain of two individuals (with $x = 0$ for $P(x)$ and 0,1 for individuals) but it is not satisfiable in a domain with only a single individual, since $P(0) \ \& \ \neg P(0)$ is false for any P.

A statement is said to be *universally valid* if it is *valid in every domain*, and satisfiable if there is a domain in which it is satisfiable; it follows that if a statement $\mathscr{A}$ is not universally valid then $\neg \mathscr{A}$ is satisfiable, and conversely, and if $\neg \mathscr{A}$ is universally valid then $\mathscr{A}$ is not satisfiable.

There are statements which are satisfiable in an infinite domain but not in any finite domain, for instance the statement

$$(\forall x)(\exists y)P(x,y) \ \& \ (\forall x) \\ \neg P(x,x) \ \& \ (\forall x)(\forall y)(\forall z)\{(P(x,y) \ \& \ P(y,z)) \rightarrow P(x,z)\}$$

(i.e. $P(x,y)$ is transitive but not reflexive and to any x corresponds a y such that $P(x,y)$). It is satisfied in the domain of the natural numbers by the predicate 'x is less than y', but not in any finite domain. For if P^* is a satisfying predicate for a finite domain then since $(\forall x)(\exists y)P^*(x,y)$ holds, to any individual i_1,

corresponds an i_2 such that $P^*(i_1,i_2)$ holds, and to i_2 corresponds i_3 such that $P^*(i_2,i_3)$ holds, and so on. Since the domain is finite the individuals i_1, i_2, i_3, . . . are not all different; let i_p, $i_{p+q}(q > 0)$ denote the same individual. Then from $P^*(i_p,i_{p+1})$, $P^*(i_{p+1},i_{p+2})$, . . ., $P^*(i_{p+q-1},i_{p+q})$ it follows that $P^*(i_p,i_{p+q})$ holds, i.e. $P^*(i_p,i_p)$ which contradicts the condition $(\forall x) \sim P^*(x,x)$.

It may readily be verified that the axioms of the predicate calculus, and any sentence derivable from them, are universally valid.

In the sentence calculus, the method of truth tables (or an equivalent method) enables us to decide of any sentence whether it is universally valid or not. For the predicate calculus this decision problem is completely solvable for a *finite* domain. Consider for instance the sentence

$$(\exists x)(\forall y)\{P(x,y) \vee \sim P(x,x)\}$$

in the domain of two individuals 0, 1. In this domain an existential statement $(\exists x)\mathscr{A}(x)$ is equivalent to the disjunction $\mathscr{A}(0) \vee \mathscr{A}(1)$, and a universal statement $(\forall x)\mathscr{A}(x)$ is equivalent to the conjunction $\mathscr{A}(0) \,\&\, \mathscr{A}(1)$. Thus the given sentence is equivalent to

$$(\forall y)\{P(0,y) \vee \sim P(0,0)\} \vee (\forall y)\{P(1,y) \vee \sim P(1,1)\}$$

and this in turn may be written as

$$[\{P(0,0) \vee \sim P(0,0)\} \,\&\, \{P(0,1) \vee \sim P(0,0)\}] \vee [\{P(1,0) \vee \sim P(1,1)\} \,\&\, \{P(1,1) \vee \sim P(1,1)\}]$$

which is valid, satisfiable, or neither according as the sentence

$$\{(p \vee \sim p) \,\&\, (q \vee \sim p)\} \vee \{(r \vee \sim s) \,\&\, (s \vee \sim s)\}$$

is universally valid, or true for certain p,q,r,s, or contradictory, and this of course is decidable by truth tables. In fact the sentence is not universally valid (since we may take q false, p true, r false, and s true) but it is true if either p and q have the same truth values, or r and s have the same truth values. Thus the original sentence is satisfiable in a domain of two individuals.

If the domain of individuals is infinite *no general decision procedure is possible*, as we shall subsequently show. There are,

however, decision procedures for a number of special forms of statements such as

$$(\forall x_1)(\forall x_2) \ldots (\forall x_p)\mathscr{A}(x_1, x_2, \ldots, x_p),$$
$$(\exists x_1)(\exists x_2) \ldots (\exists x_p)\mathscr{A}(x_1, x_2, \ldots, x_p), \text{ and}$$
$$(\forall x_1)(\forall x_2) \ldots (\forall x_p)(\exists y_1)(\exists y_2) \ldots (\exists y_q)$$
$$\mathscr{A}(x_1, x_2, \ldots, x_p, y_1, y_2, \ldots, y_q)$$

where $\mathscr{A}(\ )$ contains only free variables. In particular there is a decision procedure for any statement containing only monadic predicate variables (i.e. predicate variables with single arguments), as we shall now show.

The special feature of monadic predicates is that a sentence containing k monadic predicate variables, and no other predicate variables, is universally valid if, and only if, it is valid in a domain with 2^k elements. Let S be a sentence containing k monadic predicate variables, and let S be satisfiable in a domain D, containing more than 2^k individuals, by some choice $p_1, p_2, \ldots, p_k$ of predicates for the predicate variables $P_1, P_2, \ldots, P_k$ in S. We separate the elements of D into classes, two elements going into the same class only if the values of all the predicates $p_1, p_2, \ldots, p_k$ are the same for the two elements. Since there are at most 2^k different arrangements of truth values for k predicates, the elements of D fall into at most 2^k classes, $\alpha_1, \alpha_2, \ldots, \alpha_n$ say, $n \leq 2^k$. Denote by e_r any element of the class α_r, for $r = 1, 2, \ldots, n$, and for these values of r let us define the predicate $q_i(r)$ to have the same truth values as $p_i(e_r)$, $i = 1, 2, \ldots, k$, so that the domain of the individual variable r is the finite domain, F say, with elements $1, 2, 3, \ldots, n$. By the definition of the q_i, S is satisfiable in F by the predicates q_i if and only if S is satisfiable in D by p_i. The domain F may contain fewer than 2^k individuals but if S is satisfiable in F it will also be satisfiable in a larger domain containing exactly 2^k individuals (amongst which are the individuals of F). Thus S is universally valid if and only if it is valid in a domain of 2^k individuals.

Gödel's theorem

We have already remarked that if a sentence is provable in the predicate calculus, then it is universally valid. The converse of this is also true so that the predicate calculus is

complete. This converse is a consequence of the important theorem (Theorem G) that *every consistent set of statements is satisfiable in the domain of natural numbers.*

Let Γ be a consistent set of statements, i.e. a set from which no contradiction can be derived by predicate logic. We close all the statements of Γ by prefixing universal operators to all free individual variables after replacing each sentence containing sentence variables by the totality of sentences formed by substituting predicate variables for sentence variables. In the same way we close all the axioms of predicate logic. We denote by L the system of logic with the same elements and derivation rules as predicate logic and with the closed sentences formed from Γ and the predicate axioms for axioms. The axioms of L are consistent by hypothesis. We next add to the elements of L the individual constants 1, 2, 3, . . ., as possible values of the individual variables of L, and call the extended system L^+; the sentences of L^+ consist of the sentences of L together with all the sentences obtained by substituting individual constants for the free individual variables in predicate sentences. (In effect we form the sentences of L^+ from predicate sentences either by placing free variables within the scope of a universal quantifier or by replacing them with constants.) Further we add to L^+ a derivation rule permitting the derivation of $(\exists x)F(x)$ from $F(\mathbf{n})$, where $\mathbf{n}$ is any individual constant, and x is a variable which does not appear in $F(\mathbf{n})$, and a rule permitting the derivation of $F(\mathbf{n})$ from $(\forall x)F(x)$.

Logic L^+ is consistent, for if not we can derive both P^+ and $\neg P^+$ in L^+: the derivation of $P^+ \ \& \ \neg P^+$ contains only a finite number of formulas and so only a finite number, p say, of constants and a finite number of variables. Let $y_1, y_2, \ldots y_p$ be variables of L which do not occur in the proof of $P^+ \ \& \ \neg P^+$; replace each constant appearing in this proof by one of the y's, each constant being replaced by the same y at each point of occurrence and different constants being replaced by different variables, and let P be what P^+ becomes after this replacement. The proof of P^+ and $\neg P^+$ in L^+ becomes a proof of P and $\neg P$ in L, since the rules of inference

$$\frac{F(\mathbf{n})}{(\exists x)F(x)}, \qquad \frac{(\forall x)F(x)}{F(\mathbf{n})}$$

in L^+ are replaced by the rules

$$\frac{F(y)}{(\exists x)F(x)}, \quad \frac{(\forall x)F(x)}{F(y)}$$

in L. Thus P & $\neg$ P is derivable in L, making L inconsistent.

Suppose now that the sentences of L^+ are arranged in some order. This may be achieved for instance by assigning an order to the symbols of L^+ and then ordering sentences by the number of their symbols and lexicographically according to the order of the constituent symbols. We pick out all the sentences of the form $(\exists x)F(x)$ and suppose them enumerated as

$$(\exists x_n)F_n(x_n), \quad n = 1, 2, 3, \ldots$$

Let i_1 be the smallest integer which exceeds all the individual constants in $(\exists x_1)F_1(x_1)$ and, further, let i_{n+1} be the smallest integer which exceeds i_n and all the individual constants in $(\exists x_n)F_n(x_n)$. We extend logic L^+ by adding the axioms

$$(\exists x_r)F_r(x_r) \rightarrow F_r(i_r), \quad 1 \leq r \leq n,$$

and call the extended system L_n^+. Each L_n^+ is consistent for if not, let $k + 1$ be the first value of n for which L_n^+ is inconsistent. Then L_k^+ is consistent and L_{k+1}^+ is inconsistent ($L_0^+ = L^+$).

It follows that the addition of

$$(\exists x_{k+1})\, F_{k+1}\,(x_{k+1}) \rightarrow F_{k+1}\,(i_{k+1})$$

to L_k^+ makes the resulting system inconsistent and therefore the negation of this sentence, i.e.

$$(\exists x_{k+1})\, F_{k+1}\,(x_{k+1}) \;\&\; \neg\, F_{k+1}\,(i_{k+1})$$

is provable in L_k^+, whence it follows that

$$\{(\exists x_{k+1})\, F_{k+1}\,(x_{k+1})\} \;\&\; (\forall y) \neg\, F_{k+1}\,(y)$$

is provable in L_k^+, so that L_k^+ is inconsistent.

Let L^1 be the logic which contains all the axioms of L_n^+ for all values of n. L^1 is consistent because a derivation of a contradiction in L^1 would use only a finite number of axioms and would therefore be valid in L_n^+ for some n, which is impossible. Step by step we extend L^1 to L_1^1, L_2^1, . . . as follows.

We add to L^1 as an axiom to form L_1^1 the first sentence P such that neither P nor $\neg$ P is provable in L^1 (if any), and we

add to L_n^1 as an axiom to form L_{n+1}^1 the first sentence P such that neither P nor $\neg$ P is provable in L_n^1 (if any). Finally we take L^{11} to have all the axioms of L_n^1 for all n. For each n, L_n^1 is consistent, for if $k+1$ is the first value of n for which L_n^1 is inconsistent, then L_k^1 is consistent and if P is the axiom added to L_k^1 to form L_{k+1}^1 then it follows that $\neg$ P is provable in L_k^1, which contradicts the defining condition of L_{k+1}^1. It follows that L^{11} is consistent since a proof of a contradiction in L^{11} would be valid also in L_n^1 for some n.

We observe next that if P is any sentence in L^{11} then one of P or $\neg$ P is provable in L^{11}; for if P is the pth sentence in L^1 then there must be an $m \leq p$ for which P (or $\neg$ P) was added to L_m^1 to form L_{m+1}^1.

We come now to the crucial step in the proof.

If $F(x)$ *has no free variables other than* x *then* $(\forall x)F(x)$ *is provable in* L^{11} *if and only if all* $F(1)$, $F(2)$, $F(3)$, . . . *are provable.*

That each of $F(1)$, $F(2)$, $F(3)$, . . . is provable if $(\forall x)F(x)$ is provable follows from one of the derivation rules of L^+. For the converse we observe that if $(\forall x)F(x)$ is *not* provable in L^{11} then $(\exists x) \neg F(x)$ is provable, i.e. for a certain n, $(\exists x_n)F_n(x_n)$ is provable, whence from the axiom

$$(\exists x_n)F_n(x_n) \rightarrow F_n(i_n)$$

it follows that $F_n(i_n)$ is provable in L^{11}, i.e. $\neg F(i_n)$ is provable, which completes the proof.

The sentences of L^{11} take one of the forms

(i) $P(a_1, a_2, \ldots, a_n)$, where P is a predicate variable and $a_1, a_2, \ldots, a_n$ are integers.

(ii) $(\forall x)P(x)$, (iii) $(\exists x)P(x)$, (iv) $S \vee T$, $S \,\&\, T$, $S \rightarrow T$, $\neg S$, where S,T are themselves sentences of L^{11}.

A sentence $P(a_1, \ldots, a_n)$ is assigned the value true if and only if it is provable in L^{11}; otherwise it is assigned the value false. Since P or $\neg$ P is provable, it follows that the provable one of $P(a_1, \ldots, a_n)$, $\neg P(a_1, a_2, \ldots, a_n)$ is assigned the value true and the other the value false.

Sentences of the forms (iv) are assigned truth values in terms of the familiar interpretations of the logical connectives; thus $S \vee T$ is true if at least one of S,T is true and so on.

A sentence of the form (ii) is assigned the value true if and

only if *each* of P(1), P(2), P(3), . . . is true, so that as we have seen, $(\forall x)P(x)$ is *true* if and only if it is provable. Finally, a sentence of form (iii) is assigned the value true if and only if *one* (at least) of P(1), P(2), P(3), . . . is true, which by the nature of L^{11} occurs if and only if $(\exists x)P(x)$ is provable.

This completes the proof that L^{11} (and so *a fortiori* Γ) is satisfiable in the domain of natural numbers.

Amongst the many important conclusions which can be drawn from Theorem G we note the following:

G_1 *If* F *is not derivable in predicate logic then* $\neg$ F *is satisfiable in the domain of the natural numbers.*

For if P is the conjunction of the axioms of predicate logic, and if $\neg$ F and P were inconsistent then it follows from the deduction theorem that, for some A,

$$(P \,\&\, \neg F) \rightarrow A,$$
$$(P \,\&\, \neg F) \rightarrow \neg A$$

are both valid, whence it follows in turn that

$$(A \vee \neg A) \rightarrow \neg (P \,\&\, \neg F)$$

and

$$P \rightarrow F$$

are valid which contradicts the hypothesis that F is not derivable in predicate logic; thus $\neg$ F and P are consistent and therefore simultaneously satisfiable, by Theorem G.

G_2 *If* F *is universally valid then* F *is derivable in predicate logic,* for if F is universally valid then $\neg$ F cannot be satisfiable.

G_3 *If* F *is valid in the domain of the natural numbers then* F *is universally valid;* for F is derivable in predicate logic since otherwise $\neg$ F would be satisfiable in the domain of the natural numbers.

G_4 *If* F *is satisfiable in some non-empty domain it is satisfiable in the domain of the natural numbers;* for $\neg$ F is not universally valid, by hypothesis, and so $\neg$ F is not provable, from which it follows, by G_1, that $\neg \neg$ F, i.e. F is satisfiable in the domain of the natural numbers.

G_5 *If each finite subset of a sequence of statements* $F_1, F_2, F_3, \ldots$ *is satisfiable in some non-empty domain then the whole set is simultaneously satisfiable.*

For if the set $F_1, F_2, \ldots$ is consistent, it is satisfiable, by Theorem G; if it is inconsistent then for some A we may derive

both A and $\neg$ A from the set, and since this derivation will use only a finite number of terms of the sequence, say F_1, F_2, . . ., F_n, it follows that this finite subset is inconsistent, and therefore not satisfiable.

G_6 *If* F *is valid whenever* F_1, F_2, . . . *are all valid then* F *is derivable from* F_1, F_2, . . . *in predicate logic.*

For if F is not derivable from F_1, F_2, . . ., then $\neg$ F, F_1, F_2, . . . are consistent, and therefore simultaneously satisfiable, which contradicts the hypothesis that F is valid whenever F_1, F_2, . . . are valid.

G_7 *If the negation of every conjunction of a finite number of* F_1, F_2, . . . *cannot be proved in predicate logic then* F_1, F_2, . . . *are simultaneously satisfiable.* For if F_1, F_2, . . . are not simultaneously satisfiable, they are inconsistent and therefore, for some A, A & $\neg$ A is derivable from F_1, F_2, . . . and therefore derivable from F_1, F_2, . . ., F_n for some *n*, so that F_1, F_2, . . ., F_n is inconsistent and therefore *not* satisfiable. But since $\neg$ (F_1 & F_2 & . . . & F_n) is unprovable, therefore F_1 & F_2 & . . . & F_n is satisfiable, by G_1.

Theorems G_5 and G_6 have found important applications in modern algebra.

CHAPTER III

NUMBER THEORY

The Systems Z and Z_f. Primitive Recursive Functions. Ordinal Recursions. The Calculus of λ-conversion. Recursive Arithmetic. Arithmetical Relations.

Number Theory

The predicate calculus, like the sentence calculus, is still without any definite content. We have introduced individual variables and predicate variables but have not provided a single example of a particular predicate. This deficiency we now proceed to remedy.

To the symbols so far introduced we add the following:

The individual sign o (read zero).

The definite predicate $x = y$ (read x equals y), where x, y are individual variables.

The functions Sx (read the successor of x), $x + y$ (read x plus y) and $x \cdot y$ (read x times y).

We presuppose nothing known about these signs except that they satisfy the *axioms of arithmetic*, viz.,

E	$(x = y) \to (x = z \to y = z)$
S_1	$\neg (Sx = o)$
S_2	$(Sx = Sy) \to (x = y)$
S_3	$(x = y) \to (Sx = Sy)$
A_1	$x + o = x$
A_2	$x + Sy = S(x + y)$
P_1	$x \cdot o = o$
P_2	$x \cdot Sy = (x \cdot y) + x$

and the rule of inference I $\dfrac{A(o),\ A(x) \to A(Sx)}{A(x)}$

Axiom E is the axiom for equality. Axioms S_1, S_2 are known as the Peano axioms since they were first formulated by the Italian mathematician, G. Peano. Axioms A_1, A_2 'define' the sum function $x + y$ and P_1, P_2 'define' the product function

$x \cdot y$ (in a sense at which we shall later look more closely). Finally inference rule I is the axiom of mathematical induction.

It is easy to show that mathematical induction is not derivable from the axioms E, S_1, S_2, S_3, A_1, A_2, P_1, and P_2.

From axioms A_1 and A_2 we derive

$$0 + 0 = 0, \quad 0 + Sx = S(0 + x)$$

and from S_2

$$0 + x = x \rightarrow \{S(0 + x) = Sx\}$$

whence by sentence logic

$$0 + x = x \rightarrow (0 + Sx = Sx)$$

which, with $0 + 0 = 0$, yields by induction

$$0 + x = x.$$

Let us now add to the set of natural numbers 0, So, SSo, . . . the two new elements σ, τ satisfying the axioms:

$$S\sigma = \sigma, \quad S\tau = \tau, \quad \sigma + x = \sigma, \quad \tau + x = \tau, \quad x + \sigma = \tau, \quad x + \tau = \sigma$$

$$x \cdot \sigma = \sigma, \quad x \cdot \tau = \tau, \quad \sigma \cdot 0 = 0, \quad \tau \cdot 0 = 0, \quad \sigma \cdot x = \tau, \quad \tau \cdot x = \sigma.$$

It is readily verified that the axioms E, S, A, and P are satisfied when we permit substitution of σ and τ as well as 0 or Sx for x (or y), but induction cannot also be valid for this extended class of numbers since $0 + x = x$ is not true for the values σ, τ of x.

It is readily seen that this result remains true if we replace axiom S_3 by

$$S_3^* \qquad (\forall z) \neg (x = Sz) \rightarrow (x = 0)$$

Next we define the concepts of *term*, and *arithmetical predicate*. An individual variable is a term; 0 is a term; Sx, $x + y$, and $x \cdot y$ are terms. The result of substituting a term for a variable in a term, is a term.

The terms 0, So, SSo, . . . are called numerals.

If s, t are terms then $s = t$ is a sentence. If A and B are sentences then $\neg$ A, A & B, A $\vee$ B, A $\rightarrow$ B, $(\forall x)$A, and $(\exists x)$A are sentences. For example, $(\exists x) \neg (x = 0)$ is a sentence.

A sentence with free variables is an arithmetical predicate.

We show now that equality as defined by the equality axiom has the familiar properties expressed by the sentences $x = x$, and $(x = y) \rightarrow (y = x)$.

First we establish the inference rule

$$\frac{A(x)}{\forall x A(x)}.$$

From the provable sentence $p \to (q \to p)$, taking $A(x)$ for p and $r \vee \neg r$ for q we obtain the proved sentence

$$A(x) \to \{(r \vee \neg r) \to A(x)\}$$

whence, if $A(x)$ is provable, we derive (by *modus ponens*)

$$(r \vee \neg r) \to A(x)$$

and thence $(r \vee \neg r) \to \forall x A(x)$ by inference rule PI_1, and so, since $r \vee \neg r$ is provable,

$\forall x A(x)$ is proved, again by *modus ponens*.

By this inference rule and axiom PA_1 we obtain the further rule

$$\frac{A(x)}{A(t)}$$

which justifies the substitution of a term for a free variable in a proved sentence (or axiom).

Thus by substituting $x + 0$ for x, and x for y and for z, from axiom E we derive

$$(x + 0 = x) \to \{(x + 0 = x) \to (x = x)\}$$

whence by *modus ponens* and axiom A

$$(x + 0 = x) \to (x = x)$$

and by the same axiom and inference rule

$$x = x \text{ follows.}$$

Finally substituting x for z in axiom E, we obtain

$$(x = y) \to (x = x \to y = x)$$

whence (by sentence logic)

$$(x = x) \to \{(x = y) \to (y = x)\}$$

and so

$$(x = y) \to (y = x) \text{ is proved.}$$

The inequality $x \leqq y$ is taken to be an abbreviation for the expression $\exists z(y = x + z)$, and $x < y$ is simply the conjunction $(x \leqq y)\ \&\ \neg (x = y)$. On the basis of these definitions all the

familiar properties of inequalities are provable in our system. For instance the statements

$$(x < y) \,\&\, (y < z) \to (x < z),$$
$$x < Sx, \quad 0 < Sx, \quad 0 \leqq x,$$
$$(x < y) \vee (x = y) \vee (y < x),$$

and

$$(x < y) \to (x + z < y + z),$$

are so provable.

We can also prove that amongst the values of x for which a predicate $A(x)$ holds there is necessarily a least, i.e.

$$(\exists x)A(x) \to (\exists y)\{A(y) \,\&\, (\forall z)(z < y \to \neg A(z))\}$$

is provable, and it can be shown that this sentence is *equivalent* to mathematical induction in the sense that we employ induction to prove it, and conversely, if the existence of the least number with a given property is postulated as an axiom then we can *prove* the validity of inference by induction.

Amongst arithmetical predicates of particular interest we may mention, 'r is the remainder and q the quotient when x is divided by y', denoted by $r = \rho(x,y)$ and $q = [x/y]$ respectively; 'x is a factor of y', x/y; 'x is prime', $pr(x)$; and 'x, y are relatively prime' which is expressed by $pr(x,y)$. For non-zero values of y, $r = \rho(x,y)$ may be defined to be $(\exists q)(x = yq + r \,\&\, r < y)$; we may, however, extend the definition to include also the case $y = 0$ by taking instead the disjunction

$$(\exists q)(x = yq + r \,\&\, r < y) \vee (y = 0 \,\&\, r = x).$$

Similarly (to include the case $y = 0$) we take $q = [x/y]$ to be

$$(\exists r)(x = qy + r \,\&\, r < y) \vee (y = 0 \,\&\, q = 0)$$

On the basis of these definitions the fundamental relation

$$\{q = [x/y] \,\&\, r = \rho(x,y)\} \to (x = qy + r)$$

is provable for all values of x, y, q, and r.

The predicate x/y is of course simply $0 = \rho(y,x)$, i.e. $(\exists q)\ (y = qx)$; $pr(x)$ is

$$(\forall q)(q/x \to q = 1 \vee q = x)$$

(x and unity are the only factors of a prime x); and $pr(x,y)$ is

$$(\forall q)(q/x \,\&\, q/y \to q = 1)$$

(unity is the only common factor of a relatively prime pair x, y).

The arithmetical predicate $\omega = \rho(x,S(y \,.\, Sz))$ is of particular importance in the sequel; it can readily be shown that, for any integer p, and given relatively prime numbers d_0, d_1, d_2, . . . d_p, the sequence of $p + 1$ remainders $\rho(x,d_0)$, $\rho(x,d_1)$, $\rho(x,d_2)$, . . ., $\rho(x,d_p)$, cannot take the same run of values twice in less than $d_0 \,.\, d_1 \,.\, d_2 \,.\, .\, .\, d_p$ consecutive values of x (for if two numbers x, $x + X$ leave the same sequence of remainders when divided by d_0, d_1, . . ., d_p in turn then X is divisible by each of these numbers and so by their product). Since there are exactly $d_0 \,.\, d_1 \,.\, .\, .\, d_p$ sequences of numbers r_0, r_1, . . ., r_p with $r_0 < d_0$, $r_1 < d_1$, . . ., $r_p < d_p$, it follows that *the sequence of remainders* $\rho(x,d_0)$, $\rho(x,d_1)$, . . ., $\rho(x,d_p)$ *take any assigned run of* $p + 1$ *values once only for any* $d_0 \,.\, d_1 \,.\, .\, .\, d_p$ *consecutive values of* x. In particular given some sequence of $p + 1$ numbers v_0, v_1, . . ., v_p none of which exceeds g, and taking G to be the greater of g and p, then since the numbers $d_i = 1 + (i + 1) \,.\, 1 \,.\, 2 \,.\, 3$ . . . G, where i runs from 0 to p, are relatively prime (for the difference $d_{i+j} - d_i$ is $j \,.\, 1 \,.\, 2 \,.\, 3 \,.\, .$ G which is divisible only by the numbers 1, 2, 3, . . ., G, none of which is a factor of d) and $v_i < d_i$, for each i, it follows that there is a number x between 0 and $d_0 \,.\, d_1 \,.\, .\, .\, d_p$ such that $\rho(x,d_i) = v_i$ for each value. In other words, given *any* sequence of numbers v_0, v_1, v_2, . . ., v_p we can find numbers x and d so that $\rho(x, 1 + (i + 1)d) = v_i$* for each value of i from 0 to p.

The number system which we have been describing is known as system Z from the initial letter of the German word *Zahlentheorie*, which means number theory. System Z differs from elementary arithmetic in one important respect. In elementary arithmetic we meet a great variety of functions, but in Z we have only two, namely $x + y$ and $x \,.\, y$; such a function as x^n is lacking in Z. Of course any particular exponential, like x^3, is expressible in Z; for instance $x^3 = (x \,.\, x) \,.\, x$, but there is no means of expressing x^n with a variable n. We shall subsequently show that the relation $y = x^n$ *is* expressible in Z, so that the lack of the function x^n is less serious than it first seems, but before we show this we shall leave the system Z to consider a very important class of functions, known as *recursive functions*, a class to which $x + y$, $x \,.\, y$, and x^y all belong.

* $1 + (i + 1)d$ is another way of writing $S(d \,.\, Si)$, since 1 is an abbreviation for So.

The system of arithmetic obtained from Z by omitting axiom I (which is not a single axiom, but an infinite bundle of axioms since it is affirmed for every predicate of Z), and replacing S_3 by S_3^* is called Z_f; the importance of system Z_f, which, as we have already seen, contains only a fragment of arithmetic, lies in the fact that although it is based on a finite number of axioms it shares with Z the capacity to express (in a sense we shall later make precise) all recursive functions.

Primitive recursive functions

A function $f(t,x)$, with or without parameter t, is said to be primitive recursive in the functions $a(t)$, $b(t,x,y)$ if

$$R\begin{cases} f(t,o) = a(t) \\ f(t,Sx) = b(t,x,f(t,x)), \end{cases}$$

(where no variable occurring on the right of an equation is missing on the left though some of the variables may be absent on the right hand side). Equations R express the value of the function f for the argument Sx in terms of the value of the function for the argument x with the help of the functions a and b. The equations R are called the introductory equations of the function f, and they serve to determine, one after the other, the values of $f(t,x)$ for the values o, So, SSo, . . . of x (and any given value of the parameter t). In fact we find in turn

$$f(t,o) = a(t), f(t,So) = b(t,o,a(t)), f(t,SSo) = b(t,So,b(t,o,a(t))),$$

and so on.

Given the initial functions $Zx = o$, $Ix = x$ and the successor function Sx, a function f is said to be *primitive recursive* if there is a finite sequence of functions

$$f_1, f_2, \ldots, f_k, f$$

such that each is either an initial function or (after the second) formed by substitution from previous functions of the list, or is primitive recursive in previous functions of the list.

To express this definition in another form we may say that every primitive recursive function is either an initial function, or is obtained by substituting recursive functions in recursive functions, or is recursive in recursive functions.

The introductory equations for addition

$$a + 0 = a,\ a + Sb = S(a + b)$$

show that $a + b$ is a recursive function, being recursive in the initial functions Ix and Sx.

Similarly the product function $a \,.\, b$ introduced by the equations

$$a \,.\, 0 = 0,\ a \,.\, Sb = a \,.\, b + a$$

is recursive in Zx and $x + y$ and is therefore recursive.

If we define the exponential function a^b by

$$a^0 = 1,\ a^{Sb} = a^b \,.\, a,$$

(where we have written 1 for S0) it follows that a^b is also primitive recursive. Other primitive recursive functions we shall need are the predecessor function $P(a)$ introduced by $P(0) = 0$, $P(Sa) = a$ and the arithmetical difference $a \dot{-} b$ introduced by the equations $a \dot{-} 0 = a$, $a \dot{-} Sb = P(a \dot{-} b)$. (If $a > b$, $a \dot{-} b$ is the ordinary difference, but if $a \leqq b$ then $a \dot{-} b = 0$.) For brevity we shall sometimes denote the function $1 \dot{-} (1 \dot{-} x)$ by $\alpha(x)$, so that $\alpha(0) = 0$, $\alpha(Sx) = 1$.

We shall subsequently show that all the common functions of arithmetic are primitive recursive, but it is by no means the case that every function has this property, and in fact the sequence of functions commencing with addition, multiplication, and exponentiation and continuing by forming new functions by iteration in the same way that multiplication is obtained by repeated addition, and exponentiation by repeated multiplication, this sequence of functions, regarded as a function of 3 variables, $T(n,a,b)$, say, is known to increase more rapidly than any primitive recursive function.

The introductory equations of $T(n,a,b)$ are

$$T(0,a,b) = a \dot{+} b,\ T(1,a,0) = 0,\ T(SSn,a,0) = 1$$
$$T(Sn,a,Sb) = T(n,a,T(Sn,a,b))$$

which (apart from minor adjustments in the first row) are an instance of *doubly recursive* introductory equations. The general form of a double recursion is as follows

$$f(0,n) = \alpha(n),\ f(Sm,0) = \beta(m),$$
$$f(Sm,Sn) = \gamma(m,n,f(m,\delta(m,n,f(Sm,n))),f(Sm,n)).$$

These equations determine the value of $f(m,n)$ for any assigned m and n, for if we know the value of $f(m,n)$ for some m and *any* n, then $f(Sm,Sn)$ is given in terms of $f(Sm,n)$, whence since $f(Sm,o)$ is given, $f(Sm,n)$ is given for any n. That is, if for some m, $f(m,n)$ is determined for any n, then $f(Sm,n)$ is determined for any n, and so, since $f(o,n)$ is given for any n, it follows that $f(m,n)$ is given for any n and m. The function f may contain parameters in addition to the variables m, n and these may also be present in α, β, γ, and δ, but these latter must not contain any parameters that are not contained in f.

Just as there are doubly recursive functions which are not primitively recursive, so too trebly recursive functions may be defined which are not doubly recursive, and so on, and in fact functions may readily be defined which are not n-ply recursive for any n. It can be shown that every function which may be defined by a finite set of introductory equations from which the values of the function may be obtained by repeated substitution as in primitive and multiple recursion may also be generated using only introductory equations of the form

$$f(o) = 1,\ f(Sx) = \phi(x,\gamma(x))$$

where ϕ and γ are primitive recursive functions and $\gamma(x)$ precedes Sx in some arrangement of the natural numbers in a simple sequence (such as o 1 3 2 4 8 7 6 5 9 15 14 13 12 11 10 16 24 23 . . .). Equations of this form are known as ordinal recursions.

A predicate $P(x)$ is said to be recursive (primitive or ordinal as the case may be) if there is a recursive function $p(x)$ (primitive or ordinal), called the representing function, such that $p(x) = o$ for any x which has the property P and $p(x) = 1$ for any x which has not the property P. Similarly a predicate $P(x,y)$ is recursive if there is a recursive $p(x,y)$ which vanishes for each pair (x,y) which has the property P and takes the value unity for each pair (x,y) which has not the property, etc.

The task of identifying a primitive recursive function or relation is greatly simplified by a number of general theorems. For instance if R and S are primitive recursive relations then $\neg$ R, R $\vee$ S, R & S, R $\rightarrow$ S are primitive recursive relations, with representing functions $1 \dot{-} r$, $r \cdot s$, $(r + s) \dot{-} rs$, $(1 \dot{-} r)s$ where r and s are the representing functions of R and S: for

$1 \dot{-} r$ takes the value 0 only if $r = 1$ and the value 1 otherwise, $r \cdot s$ vanishes if either r or s vanishes (either R or S is true), $(r + s) \dot{-} rs$ vanishes only if both r and s vanish (both R and S are true) and $(1 \dot{-} r)s = 1$ only if $r = 0$ and $s = 1$.

If f and g are primitive recursive functions then

$$f \leqq g, \quad f = g, \quad f > g$$

are primitive recursive relations, for the representing function of $f \leqq g$ is $\alpha(f \dot{-} g)$, and $f = g$, $f > g$ are equivalent to $(f \leqq g)\ \&\ (g \leqq f)$, $\neg (f \leqq g)$ respectively.

Another important set of results are the *bounded quantifier* theorems: If the predicate $R(x,y)$ is primitive recursive then the relations E, A where

$$E(y,z) = (\exists x)\{x \leqq z\ \&\ R(x,y)\},$$
$$A(y,z) = (\forall x)\{x \leqq z \to R(x,y)\},$$

and the function $(\mu x)(x \leqq z\ \&\ R(x,y))$ denoting the least value of x, not exceeding z, such that $R(x,y)$ holds, when there is such an x, and taking the value zero, otherwise, are all primitive recursive.

The representing function of E, for instance is $\prod_{x=0}^{z} r(x,y)$ where $r(x,y)$ is the representing function of $R(x,y)$ and $\prod_{x=0}^{0} r(x,y) = r(0,y)$, $\prod_{x=0}^{Sz} r(x,y) = \{\prod_{x=0}^{z} r(x,y)\} \cdot r(Sz,y)$ and that of A is $1 \dot{-} \{1 \dot{-} \sum_{x=0}^{z} r(x,y)\}$ where

$$\sum_{x=0}^{0} r(x,y) = r(0,y), \quad \sum_{x=0}^{Sz} r(x,y) = \sum_{x=0}^{z} r(x,y) + r(Sz,y).$$

As examples of the bounded operator theorems we notice that the following functions and relations are primitive recursive:

'The whole part of a divided by b', which may be defined as

$$[a/b] = (\mu x)\{x \leqq a\ \&\ Sa \leqq b \cdot Sx\};$$

'the whole part of root n'

$$[n^{\frac{1}{2}}] = (\mu x)\{x \leqq n\ \&\ (Sx)^2 > n\}$$

'a divides b',

$$a/b = (\exists x)\{x \leqq b\ \&\ ax = b\};$$

'p is prime',

$$\Pr p \rightleftharpoons (p > 1) \;\&\; (\forall x)[x \leqq p \to \{x/p\} \to (x = \mathrm{S}0) \vee (x = p)\}];$$

'p_n is the $(n+1)$th prime',

$$p_0 = \mathrm{SS}0,\ p_{\mathrm{S}n} = (\mu x)\{x \leqq p_n ! + 1 \;\&\; x > p_n \;\&\; \Pr x\};$$

and finally 'the exponent of the greatest power of p_n which divides a' may be defined as

$$\{a\}_n = (\mu x)\{p_n^x | a \;\&\; \neg\, p_n^{\mathrm{S}x} | a\}.$$

Course-of-values recursion

There are many kinds of recursive definition which, though seemingly very different from primitive recursion, may in fact be transformed into a primitive recursion.

Consider, for instance, the sequence $f(0), f(1), f(2), \ldots$ in which $f(0) = a$, $f(1) = b$, $f(2) = f(1) + f(0) = a + b$, $f(3) = f(2) + f(1) = a + 2b$, and so on. The general law of the sequence is $f(n+1) = f(n) + f(n-1)$, so that $f(n+1)$ depends, not just on $f(n)$ but also on $f(n-1)$. To show that this sequence $f(n)$ may also be obtained by primitive recursions alone, we introduce the function

$$g(n) = \prod_{r=0}^{n} p_r^{f(r)}$$

so that $f(n) = \{g(n)\}_n$, the exponent of the greatest power of the prime p_n which divides $g(n)$.

Let $\gamma(n,b) = \{b\}_n + \{b\}_{n-1}$, so that

$$\gamma(n, g(n)) = f(n) + f(n \dot{-} 1) = f(n+1).$$

It follows that

$$g(n+1) = p_{n+1}^{f(n+1)} \cdot g(n) = p_{n+1}^{\gamma(n,g(n))} \cdot g(n)$$

so that $g(n)$ is primitive recursive and therefore $f(n)$ is primitive recursive.

The function $\prod_{r=0}^{n} p_r^{f(r)}$ is called the course-of-values function of $f(n)$, and a recursion in which $f(n+1)$ depends not just on $f(n)$ but also on the values of $f(x)$ for values of $x < n+1$, is called a course-of-values recursion. The foregoing method of

transforming a course-of-values recursion into a primitive recursion is of general applicability.

Recursion with parameter substitution

Another recursion transformable to primitive recursion is recursion with parameter substitution. As an example of such a recursion we consider

$$f(0,a) = a,\ f(n+1,a) = f(n,\gamma(n,a)).$$

To determine $f(n+1,a)$ from the second of these equations we need to know the value of $f(n,x)$ not just for $x = a$, but for the value $\gamma(n,a)$ of x, which of course varies with n.

The method of transforming this recursion is of considerable interest apart from the present application. If we calculate in turn the values of $f(n,a)$ for $n = 0, 1, 2, 3$, and so on, we determine the sequence of terms

$$a,\ \gamma(0,a),\ \gamma(0,\gamma(1,a)),\ \gamma(0,\gamma(1,\gamma(2,a))),$$

and so on, which are formed by repeated substitution for the parameter. The essential idea of the transformation is to disentangle these substitutions by means of a function $\psi(n,a)$ with the following property:

For any n we can find p,q and for any p,q we can find n so that

$$\psi(n+1,a) = \gamma(p,\psi(q,a)).$$

With a suitable initial condition, like $\psi(0,a) = a$, this function transforms any term like $\gamma(0,\gamma(1,\gamma(2,a)))$ successively into $\gamma(0,\gamma(1,\gamma(2,\psi(0,a))))$, $\gamma(0,\gamma(1,\psi(h_1,a)))$, $\gamma(0,\psi(h_2,a))$, and finally into $\psi(h_3,a)$, for appropriate h_1, h_2, h_3, so that if ψ can be defined by primitive recursion, and also the auxiliary function h_r, then so can $f(n,a)$.

To construct such a function $\psi(n,a)$ we observe that, for any n, $n+1$ is expressible in only one way in the form $2^p(2q+1)$, where in fact $p = \{n+1\}_0$ and $q = [(n+1)/2^{p+1}]$, so that p and q are primitive recursive functions of n (and of course n is primitive recursive in p and q). We define

$$\psi(o,a) = a,\ \psi(n+1,a) = \gamma(p,\psi(q,a))$$

(where p, q are the functions of n just introduced); since $q < n+1$ this is a course-of-values recursion so that ψ is

primitive recursive. To complete the transformation it remains to show that there is a primitive recursive $k(n)$ such that $f(n,a) = \psi(k(n),a)$; first, however, we introduce a $g(n,a)$ such that

$$\psi(n,\psi(s,a)) = \psi(g(n,s),a).$$

To determine $g(n,a)$ we consider the relation

$$\begin{aligned}\psi(g(n+1,s),a) &= \psi(n+1,\psi(s,a))\\ &= \gamma(p,\psi(q,\psi(s,a)))\\ &= \gamma(p,\psi(g(q,s),a))\\ &= \psi(2^p(2g(q,s)+1),a)\end{aligned}$$

which reveals the definition by course-of-values recursion

$$g(n+1,s) = 2^{p+1}\cdot g(q,s) + 2^p,$$

the definition of g being completed by taking $g(0,s) = s$. Similarly, to determine $k(n)$ we consider

$$\begin{aligned}\psi(k(n+1),a) &= f(n+1,a) = f(n,\gamma(n,a)) = \psi(k(n),\gamma(n,a))\\ &= \psi(k(n),\psi(2^n,a)) = \psi(g(k(n),2^n),a)\end{aligned}$$

whence $k(n+1) = g(k(n),2^n)$, which together with the initial condition $k(0) = 0$ is a primitive recursive definition of the function $k(n)$.

Simultaneous recursion

As a final illustration of an indirect definition of a primitive recursive function we consider the *simultaneous recursion*

$$\begin{aligned}f(0) &= g(0) = 0\\ f(n+1) &= \mathrm{P}(f(n),g(n))\\ g(n+1) &= \mathrm{Q}(f(n),g(n))\end{aligned}$$

where P, Q are primitive recursive.

Here again we introduce an auxiliary function

$$h(n) = 2^{f(n)}\cdot 3^{g(n)}$$

so that $f(n) = \{h(n)\}_0,\ g(n) = \{h(n)\}_1.$

It remains to show that $h(n)$ is primitive recursive.

Writing $p(x) = \mathrm{P}(\{x\}_0,\{x\}_1),\ q(x) = \mathrm{Q}(\{x\}_0,\{x\}_1)$

and $\mu(x) = 2^{p(x)} \cdot 3^{q(x)}$,

we have

$$f(n+1) = \mathrm{P}(f(n),g(n)) = p(h(n))$$
$$g(n+1) = \mathrm{Q}(f(n),g(n)) = q(h(n))$$

and so

$$h(n+1) = 2^{f(n-1)} \cdot 3^{g(n+1)} = 2^{p(h(n))} \cdot 3^{q(h(n))} = \mu(h(n))$$

which, together with the initial condition $h(0) = 1$, completes the primitive recursive definition of $h(n)$.

Canonical forms

We conclude this section on recursion by noting a simple canonical form for primitive recursive functions. It is known that all primitive recursive functions *of one variable* may be obtained from the initial functions

$$\mathrm{S}n, \quad n \dot{-} [n^{\frac{1}{2}}]^2$$

using only the operations

$$\alpha(n) + \beta(n), \quad \beta(\alpha(n)), \quad \beta^n 0$$

to form new functions from the initial functions or from functions so obtained, where $\beta^n 0$ denotes the function $f(n)$ defined by the primitive recursion $f(0) = 0, f(\mathrm{S}n) = \beta(f(n))$.

All primitive recursive functions of more than one variable can be obtained from primitive recursive functions of one variable and the function $x \cdot y$, by substitution alone. For if $\mathrm{F}(u,v)$ is a two variable primitive recursive function, then

$$f(x) = \mathrm{F}(\{x\}_0,\{x\}_1)$$

is a one-variable primitive recursive function; but

$$\mathrm{F}(u,v) = f(2^u \cdot 3^v)$$

and so $\mathrm{F}(u,v)$ may be obtained by substitution alone from the functions

$$f(x), 2^u, 3^v, x \cdot y.$$

In the case of a function of more than two variables, the same device serves to reduce the variables one by one.

There is a remarkably similar canonical form for ordinal

recursive functions. All one variable ordinal recursive functions may be obtained from the initial functions

$$Sn,\ n \dot{-} [n^{\frac{1}{2}}]^2$$

using only the operations

$$\alpha(n) + \beta(n),\ \beta(\alpha(x)),\ \beta^{-1}(x),$$

the last function, denoting the *inverse* of $\beta(x)$, being used only when the values of $\beta(x)$ include all natural numbers.

Exactly as for primitive recursive functions, ordinal recursive functions of more than one variable can be obtained from ordinal recursive functions of one variable and the function $x \cdot y$, by substitution alone; the foregoing proof for primitive recursion applies unchanged to the case of ordinal recursive functions.

The Conversion Calculus

A canonical form for recursive functions is provided by the Church calculus of λ-conversion. Expressions are formed from a class of variables $x,y,z, \ldots, f,g,h, \ldots$ by the two operations of concatenation and abstraction. Concatenation is simply the operation of forming the expression "(AB)" from the expressions "A" and "B", and abstraction forms "$\lambda x|$M" from "M". Thus for instance from f and x we may form, in turn, by concatenation (fx), $(f(fx))$, $(f(f(fx)))$ and so on. From fx by abstraction we form $\lambda x|fx$ and thence $\lambda f|\lambda x|fx$.

The formula $\lambda x_1|\lambda x_2|\lambda x_3| \ldots |\lambda x_n|$M

is abbreviated to $\lambda x_1 x_2 x_3 \ldots x_n|$M.

The abstraction $\lambda x|$M may be interpreted as the function of x for which M holds (where M is supposed to contain the variable x) and concatenation AB is (in effect) providing a value B for the variable in A. This intended interpretation explains the transformation rule by which $(\lambda x|\text{M})$N may be transformed (or converted) into the expression which results by substituting "N" for all occurrences of "x" in "M". Thus for instance $(\lambda x|x)$N is convertible to N, and

$$((\lambda fx|f(f(fx)))\text{G})\text{A},$$

i.e.

$$((\lambda f|\lambda x|f(f(fx)))\text{G})\text{A}$$

is convertible first to

$$(\lambda x|\mathrm{G}(\mathrm{G}(\mathrm{G}x)))\mathrm{A}$$

and then to $\mathrm{G}(\mathrm{G}(\mathrm{GA}))$.

Thus $((\lambda fx|f(f(fx)))\mathrm{G})\mathrm{A}$ may be transformed into the expression which results when we substitute G for f and A for x in the expression $f(f(fx))$.

If M contains occurrences of x but not of y and if the result of substituting y for x in M is N, then, for any A, $(\lambda x|\mathrm{M})\mathrm{A}$ and $(\lambda y|\mathrm{N})\mathrm{A}$ are convertible into the same expression, namely the expression which results from the substitution of A for x in M (or for y in N). For this reason we say that $\lambda x|\mathrm{M}$ is convertible into $\lambda y|\mathrm{N}$. Denoting by $\mathrm{S}^x_y\mathrm{M}$ the expression which is obtained by substituting y for the variable x in M (or M itself if M does not contain x), a more rigorous statement of the rules of conversion may be given as follows:

C_1. Any part of M may be replaced by $\mathrm{S}^x_y\mathrm{M}$ provided that x is *not* free in M and y does *not* occur in M.

C_2. $(\lambda x|\mathrm{M})\mathrm{N}$ may be replaced by $\mathrm{S}^x_N\mathrm{M}$ provided that the bound variables of M are distinct both from x and from the free variables of N.

C_3. $\mathrm{S}^x_N\mathrm{M}$ may be replaced by $(\lambda x|\mathrm{M})\mathrm{N}$ in every context in which the converse replacement is permitted by C_2.

If $\mathrm{M}_0, \mathrm{M}_1, \mathrm{M}_2, \ldots, \mathrm{M}_k, \mathrm{M}_{k+1}$ is a sequence of expressions such that for each r, $0 \leq r \leq k$, M_r may be converted into M_{r+1} by one of the rules, C_1, C_2, C_3 then M_0 and M_{k+1} are said to be interconvertible (or simply equal). Interconvertibility is transitive (by definition), symmetric because rules C_2 and C_3 are converses and because C_1 is its own converse, and therefore also reflexive.

We do not attempt to λ-define 0, but 1 is taken to be λ-defined by the expression $\lambda fx|(fx)$.

The successor function S is λ-defined to be

$$\lambda nfx|f((nf)x).$$

It follows that the numeral S1 is equal to

$$\lambda fx|f((1f)x),$$

and since $1f$ is equal to

$$\lambda x|fx$$

therefore 2 (= S1) is equal to

$$\lambda fx|f(fx);$$

similarly 3 = S(S1) is equal to

$$\lambda fx|f((2f)x)$$

and since $(2f)x$ is equal to $f(fx)$ we find 3 is equal to

$$\lambda fx|(f(f(fx)))$$

and so on.

We have seen that $(1f)x$ is equal to fx and $(2f)x$ to $f(f(x))$; it is easily seen that for any numeral m, $(mf)x$ is equal to the expression $f(f \ldots f(fx) \ldots)$ with mf's. For this is true with $m = 1$, and if it is true for some m, then $(Smf)x$ is equal to $f((mf)x)$ and therefore equal to the expression $f(f(f \ldots f(fx) \ldots))$ with $m + 1$ f's.

A function F defined for all natural number arguments is said to be *λ-definable* if there is an expression of the conversion calculus **F** such that for all natural numbers m,n satisfying $Fm = n$, we have $\mathbf{FM} = \mathrm{N}$ where M and N represent the numbers m and n in the calculus; similarly for functions F of two arguments, F is λ-definable if $Flm = n$ entails $\mathbf{FLM} = \mathrm{N}$ and so on.

The sum function $m + n$ is λ-definable by the expression

$$\lambda fx|(mf(nfx)).$$

To prove this we remark first that

$$\lambda fx|mf(1fx) \text{ is equal to } \lambda fx|mf(fx)$$

which in turn is equal to Sm since $mf(fx)$ and $f((mf)x)$ are both equal to the expression $f(f(\ldots f(x) \ldots))$ with $m + 1$ f's. Next we observe that if $\lambda fx|(mf(nfx))$ does equal $m + n$ for some m and n then

$$\lambda fx|(mf(Snfx))$$

is equal to

$$\lambda fx|mf(f((nf)x))$$

i.e. to

$$\lambda fx|f(mf((nf)x));$$

but $S(m + n)$ is equal to

$$\lambda fx|f(\overline{(m + n}f)x)$$

which by hypothesis equals

$$\lambda fx|f(mf(nfx)),$$

completing an inductive proof.

The product function $m \cdot n$ is λ-definable by the expression

$$\lambda f|(m(nf)).$$

We start by proving that the expression

$$(nf)((nf)((nf)(\ .\ .\ .(nfx))\ .\ .\ .)),$$

containing (nf) m-times, is equal to $(m \cdot n)fx$.

With $m = 1$ the result is evident, and if it is true for some m, then

$$(nf)((nf)((nf)\ .\ .\ .(nfx)\ .\ .\ .)),$$

containing (nf) repeated $(m + 1)$ times is equal to

$$(nf)((m \cdot n)fx)$$

which in turn is equal to $(n + m \cdot n)fx$, and so to $((m + 1) \cdot n)fx$ completing the inductive proof.

It follows that

$$\begin{aligned} &\lambda f|m(nf) \\ &= \lambda f|\lambda x|nf(nf(\ .\ .\ .(nfx)\ .\ .\ .)), \text{ with } nf \text{ occurring } m \text{ times,} \\ &= \lambda fx|(m \cdot n)fx = (m \cdot n), \end{aligned}$$

for any assigned integers m and n.

The exponential function m^n admits the particularly simple definition (nm).

For $(1m) = \lambda x|mx = \lambda f|mf = \lambda fx|f(f(\ .\ .\ .(fx)\ .\ .\ .))$ with m iterations, $= m$,

and if $(nm) = m^n$ for a certain n then

$$\begin{aligned} ((Sn)m) &= \lambda x|m(nm)x) \\ &= \lambda f|m(m^n f), \text{ by hypothesis,} \\ &= m \cdot m^n = m^{n+1}. \end{aligned}$$

It may be shown that every primitive recursive function is λ-definable, and in fact that every recursive function is so definable.

Recursive arithmetic

The part of arithmetic which treats only of recursive functions and relations is called recursive arithmetic. Recursive

arithmetic comprises all the familiar theorems of the classical theory of numbers as is apparent from the range of functions and relations which we have shown to be recursive. Recursive arithmetic is, however, more limited in resources than the number system Z for it is known that there is a primitive recursive predicate $T(x,y)$ such that the predicate $(\exists y)T(x,y)$ is *not* recursive.

Recursive arithmetic may be based upon the axioms of sentence logic (postulated for statements involving only recursive relations) and the axioms of arithmetic (p. 44) fortified by the introductory equations for every recursive function, postulated as axioms. The additional axioms of predicate logic must be excluded since they introduce the operators $\forall$ and $\exists$, and as we have already remarked, the relation $(\exists y)R(x,y)$ may fail to be recursive even when $R(x,y)$ is primitive recursive. Thus recursive arithmetic is a free variable formal system without quantifiers. We are going to show now that *recursive arithmetic may in fact be set up without the aid of any logical axioms whatever*, in a calculus of equations between functions in which EVERY tautology is provable.

In the equation calculus every sentence takes the form $F = G$ where F and G are recursive functions (of any number of variables). The axioms are simply explicit definitions and the introductory equations of recursive functions. For derivation schemes we have the *equalizing rule* (E)

$$\begin{array}{l} F(0) = G(0) \\ F(Sx) = H(x,F(x)) \\ \underline{G(Sx) = H(x,G(x))} \\ F(x) = G(x) \end{array}$$

(which says, in effect, that primitive recursive introductory equations define a function uniquely), and the substitution rules

$$(Sb_1)\ \frac{F(x) = G(x)}{F(A) = G(A)} \qquad (Sb_2)\ \frac{A = B}{F(A) = F(B)}$$

$$(T)\ \begin{array}{l} F = G \\ \underline{F = H} \\ G = H \end{array}$$

A proof in the calculus is a finite sequence of equations each of which is either an axiom or a proved equation or follows from one or more previous equations of the proof by one of the derivation schemes; the last equation of a proof is said to be proved by the proof.

To prove $A = A$ for any term A, we note that the three equations $x + 0 = x$, $x + 0 = x$, $x = x$ form a proof (the first two being introductory equations for the sum function and the third following from them by the substitution rule T); finally the equation $A = A$ is derived from $x = x$ by the substitution rule Sb_1. It follows that $B = A$ is derivable from $A = B$, since it is derivable from $A = B$, $A = A$ by the substitution rule T.

To illustrate the operation of the calculus we prove some of the familiar properties of the sum, difference, and product functions. We start with the commutative property of addition

$$x + y = y + x.$$

The proof turns on the following analogues of the introductory equations for addition (axioms A_1, A_2):

$$(A_1') \quad 0 + x = x, \qquad (A_2') \quad Sy + x = S(y + x).$$

The first of these is a consequence, by the equalizing rule, of the three equations

$$0 + 0 = 0, \quad 0 + Sx = S(0 + x), \quad Sx = Sx$$

(the first two following from axioms A_1, A_2 by the substitution rules) with the function St taking the place of $H(x,t)$ in the application of the equalizing rule. The proof of the second equation runs as follows

(a) $x + 0 = x$	(b) $Sy + 0 = Sy$
(c) $S(y + 0) = Sy$	(d) $Sy + 0 = S(y + 0)$
(e) $x + Sy = S(x + y)$	(f) $Sy + Sx = S(Sy + x)$
(g) $y + Sx = S(y + x)$	(h) $S(y + Sx) = SS(y + x)$
(i) $Sy + x = S(y + x)$	

In this proof (a) is an axiom; (b) follows from (a) by substituting Sy for x, and (c) from (a) by substituting y for x and using Sb_2, and (d) follows from (b) and (c) by T. In the next line, (e) is an axiom from which (f) and (g) follow by four applications of Sb_1; (h) follows from (g) by Sb_2 and finally (i) follows from

(d), (f), (h) by the equalizing rule E, with $Sy + x$, $S(y + x)$ as the values of $F(x)$, $G(x)$ and St as the value of $H(x,t)$, y playing the part of a parameter concealed in F and G.

From A_1 and A_1' follows $x + 0 = 0 + x$, by T, and from this equation together with A_2 and A_2' we derive $x + y = y + x$ again by E (with y in place of x, and x playing the part of a parameter).

The following proofs we give in outline only, running a number of equations together for brevity.

Theorem $(a \dot{+} b) + c = a + (b \dot{+} c)$

Proof
$$(a \dot{+} b) + 0 = (a + b) = a + (b + 0)$$
$$(a \dot{+} b) + Sc = S((a + b) \dot{+} c)$$
$$a + (b \dot{+} Sc) = a + S(b \dot{+} c) = S(a + (b + c)).$$

It follows that

$$(a \dot{+} b) + c = a + (b \dot{+} c) = a + (c + b)$$
$$= (a + c) + b.$$

Theorem $Sa \,.\, b = (a \,.\, b) + b$

Proof
$$(Sa) \,.\, 0 = 0 = a \,.\, 0 + 0$$
$$Sa \,.\, Sb = (Sa) \,.\, b + Sa$$
$$a \,.\, Sb + Sb = (a \,.\, b + a) + Sb$$
$$= S((a \,.\, b + a) + b)$$
$$= S((a \,.\, b + b) + a)$$
$$= (a \,.\, b + b) + Sa.$$

Theorem $(a \dot{-} b) \dot{-} 1 = (a \dot{-} 1) \dot{-} b$

Proof
$$(a \dot{-} 0) \dot{-} 1 = (a \dot{-} 1) \dot{-} 0$$
$$(a \dot{-} Sb) \dot{-} 1 = ((a \dot{-} b) \dot{-} 1) \dot{-} 1$$
$$(a \dot{-} 1) \dot{-} Sb = ((a \dot{-} 1) \dot{-} b) \dot{-} 1$$

It follows that

$$Sa \dot{-} Sb = (Sa \dot{-} b) \dot{-} 1 = (Sa \dot{-} 1) \dot{-} b$$
$$= a \dot{-} b$$

We mention next two particular cases of the equalizing rule E:

$$(Eq_1)\ \frac{F(Sx) = F(x)}{F(x) = F(0)}, \qquad (Eq_2)\ \frac{F(0) = 0,\ F(Sx) = 0}{F(x) = 0}.$$

To prove Eq_1 we introduce $H_1(x,t)$, $C(t)$ by the explicit definitions

$$H_1(x,t) = t, \quad C(t) = F(0)$$

whence we derive $C(0) = F(0)$ and $C(Sx) = H_1(x,C(x))$; but from $F(Sx) = F(x)$ follows $F(Sx) = H_1(x,F(x))$, and thence, by E, we derive $F(x) = C(x)$, and finally, by T, we reach $F(x) = F(0)$. To prove Eq_2 we introduce $Z(t)$ by the explicit definition $Z(t) = 0$, whence, by substitution, $Z(F(x)) = 0$, and from $F(Sx) = 0$ we derive

$$F(Sx) = Z(F(x));$$

but $$Z(Sx) = Z(Z(x))$$

and $$F(0) = Z(0)$$

and so, by E, $$F(x) = Z(x) = 0.$$

As an illustration of the schema Eq_1 we prove $0 \cdot a = 0$; for $0 \cdot Sa = 0 \cdot a + 0 = 0 \cdot a$, whence $0 \cdot a = 0 \cdot 0 = 0$.

Similarly

$$0 \dot{-} x = 0,$$

for $$0 \dot{-} Sx = (0 \dot{-} x) \dot{-} 1 = (0 \dot{-} 1) \dot{-} x = 0 \dot{-} x,$$

and so $$0 \dot{-} x = 0 \dot{-} 0 = 0;$$

and $$(a + b) \dot{-} b = a,$$

for

$$(a + Sb) \dot{-} Sb = S(a + b) \dot{-} Sb = (a + b) \dot{-} b,$$

so that

$$(a + b) \dot{-} b = (a + 0) \dot{-} 0 = a.$$

From the two proved equations $(a + b) \dot{-} b = a$ and $0 \dot{-} a = 0$ we derive the schema $\dfrac{F + G = 0}{F = 0}$; for from $F + G = 0$ follows $(F + G) \dot{-} G = (0 \dot{-} G)$ and thence $F = 0$.

By means of Eq_2 we prove the very important equation

$$x(1 \dot{-} x) = 0;$$

for $$0(1 \dot{-} 0) = 0$$

and $$Sx(1 \dot{-} Sx) = Sx(So \dot{-} Sx) = Sx(0 \dot{-} x) = 0.$$

We conclude this section by proving the theorems

$$a \cdot b = b \cdot a,\ a(b + c) = a \cdot b + a \cdot c,$$
$$a \dot{-} (b + c) = (a \dot{-} b) \dot{-} c$$

We have proved

$$a \cdot 0 = 0 \cdot a,\ a \cdot Sb = a \cdot b + a,\ Sb \cdot a = b \cdot a + a$$

from which $a \cdot b = b \cdot a$ follows. The proof of the second equation consists in

$$a(b \dot{-} 0) = a \cdot b = a \cdot b + a \cdot 0$$
$$a(b + Sc) = a \cdot S(b + c) = a(b + c) \dot{+} a$$
$$a \cdot b + a \cdot Sc = a \cdot b + (a \cdot c \dot{+} a) = (a \cdot b + a \cdot c) \dot{+} a$$

For the third we have

$$a \dot{-} (b \dot{-} 0) = (a \dot{-} b) \dot{-} 0$$
$$a \dot{-} (b + Sc) = (a \dot{-} (b + c)) \dot{-} 1$$

and
$$(a \dot{-} b) \dot{-} Sc = ((a \dot{-} b) \dot{-} c) \dot{-} 1$$

We list now, without further proof, a few further key theorems.

The first of these is

$$a + (b \dot{-} a) = b \dot{+} (a \dot{-} b).$$

Writing $|a,b|$ for the positive difference between a and b, i.e. $(a \dot{-} b) \dot{+} (b \dot{-} a)$, this equation serves to pass from

$$|F,G| = 0$$

to
$$F = G;$$

for from $|F,G| = 0$ we derive both $F \dot{-} G = 0$ and $G \dot{-} F = 0$ and thence, from $F \dot{+} (G \dot{-} F) = G + (F \dot{-} G)$, follows

$$F = G.$$

The converse derivation of $|F,G| = 0$ from $F = G$ is trivial. This result serves to derive from Eq_2 the more general scheme

$$\frac{f(0) = g(0)}{f(Sx) = g(Sx)}$$
$$f(x) = g(x);$$

for if $\phi(x) = |f(x), g(x)|$ then from $f(0) = g(0)$ and $f(Sx) = g(Sx)$ we derive $\phi(0) = 0$, $\phi(Sx) = 0$, whence by Eq_2, $\phi(x) = 0$ and so finally $f(x) = g(x)$.

As an example of this schema we prove $a(x \dot{-} 1) = ax \dot{-} a$. We have

$$a(0 \dot{-} 1) = 0 = a \,.\, 0 \dot{-} a$$
$$a(Sx \dot{-} 1) = ax$$
$$a \,.\, Sx \dot{-} a = (ax + a) \dot{-} a = ax,$$

which completes the proof.

It follows that $a(b \dot{-} c) = a \,.\, b \dot{-} a \,.\, c$

for
$$a(b \dot{-} 0) = a \,.\, b$$
$$a \,.\, b \dot{-} a \,.\, 0 = a \,.\, b$$
$$a(b \dot{-} Sc) = a(b \dot{-} c) \dot{-} a$$

and
$$a \,.\, b \dot{-} a \,.\, Sc = (a \,.\, b \dot{-} a \,.\, c) \dot{-} a.$$

Next we notice the schema

$$\frac{F(0) = 0,\ (1 \dot{-} F(n))F(Sn) = 0}{F(n) = 0}$$

which plays the part of mathematical induction, and the *substitution formula*

$$(1 \dot{-} |x,y|)F(x) = (1 \dot{-} |x,y|)F(y).$$

We conclude this section by proving the two-variable schema

$$\frac{f(a,0) = 0,\ f(0,b) = 0,\ f(Sa,Sb) = f(a,b)}{f(a,b) = 0}.$$

From the two equations

$$f(Sa,0) = f(a,0),\ f(Sa,Sb) = f(a,b)$$

follows

$$f(Sa,b) = f(a,b \dot{-} 1)$$

which, together with

$$f(0,b) = f(0,b \dot{-} 1)$$

proves

$$f(a,b) = f(a \dot{-} 1,b \dot{-} 1)$$

whence, by substitution,

$$f(a \dot{-} n,b \dot{-} n) = f(a \dot{-} Sn,b \dot{-} Sn).$$

From this, it follows by Eq_1 that

$$f(a \dot{-} n,b \dot{-} n) = f(a,b)$$

and so, substituting b for n,

$$f(a,b) = 0.$$

By considering the difference $|f(a,b),g(a,b)|$ we immediately deduce the schema

$$\frac{f(a,0) = g(a,0), f(0,b) = g(0,b)}{f(Sa,Sb) = f(a,b),\ g(Sa,Sb) = g(a,b)} \cdot \qquad f(a,b) = g(a,b)$$

As an example of the use of this schema we may mention the equation (used in the sequel)

$$1 \dot{-} (a \dot{-} b) = \{1 \dot{-} (Sa \dot{-} b)\} + \{1 \dot{-} |a,b|\}.$$

Each side of the equation is unchanged by writing Sa,Sb for a,b, and the equation is obviously true when either a or b is zero.

The logical connectives &, $\vee$, $\rightharpoonup$ and $\to$ between recursive relations may be introduced into the equation calculus as *standing for* the corresponding representing equations. Thus we may introduce $\rightharpoonup (F = 0)$ for $1 \dot{-} F = 0$; $(F = 0)$ & $(G = 0)$ to stand for the equation $F + G = 0$; $(F = 0) \vee (G = 0)$ for $F \cdot G = 0$; and $(F = 0) \to (G = 0)$ for $(1 \dot{-} F)G = 0$. In particular if we associate with atomic statements p, q, r, etc., variables p, q, r, etc., the value zero of the associated variable corresponding to the truth of the proposition (so that a proposition p is asserted by the equation $p = 0$ and denied by the equation $1 \dot{-} p = 0$), then it becomes possible to express in the equation calculus any statement of the sentence calculus, and any statement of the predicate calculus which does not introduce the universal and existential operators. *All the axioms of the predicate calculus, as well as the axioms of the number system Z (apart from the defining equations for the sum and product functions) become* provable *equations in the calculus.* We consider first the four axioms

(i) $p \vee p \to p$, (ii) $p \to p \vee q$,

(iii) $p \vee q \to q \vee p$, (iv) $(p \to q) \to (r \vee p \to r \vee q)$.

Axiom (i) is expressed in the calculus by the equation

$$(1 \dot{-} p \cdot p)p = 0.$$

From the proved equation $(1 \dot{-} x)x = 0$ we obtain both

$$(1 \dot{-} p \cdot p)p \cdot p = 0 \text{ and } (1 \dot{-} p \cdot p)p(1 \dot{-} p) = 0$$

whence by addition

$$(1 \dot{-} p \cdot p)p(p + (1 \dot{-} p)) = 0$$

and so $$(1 \dot{-} p \cdot p)p(1 + (p \dot{-} 1)) = 0$$

from which we prove

$$(1 \dot{-} p \cdot p)p = 0.$$

Axiom (ii) is expressed by

$$(1 \dot{-} p)p \cdot q = 0$$

which follows immediately from the proved equation $(1 \dot{-} p)p = 0$. The third axiom's representation is

$$(1 \dot{-} p \cdot q)q \cdot p = 0$$

which follows from the commutative property of multiplication, and the fourth axiom appears as

$$(1 \dot{-} (1 \dot{-} p)q)(1 \dot{-} r \cdot p)r \cdot q = 0;$$

writing $f(p)$ for the left hand side of this equation, we have

$$f(0) = (1 \dot{-} q)q \cdot r = 0$$

$$f(Sp) = \{1 \dot{-} r \cdot (p + 1)\}r \cdot q = ((1 \dot{-} r) \dot{-} r \cdot p)r \cdot q$$

$$= ((1 \dot{-} r)r \dot{-} r \cdot r \cdot p)q = (0 \dot{-} r \cdot r \cdot p)q = 0$$

which proves $f(p) = 0$, by Eq_2.

There remain to be proved the axioms E, S_1, S_2, S_3 and I of number system Z.

For E we require to prove

$$(1 \dot{-} |x,y|)(1 \dot{-} |x,z|)|y,z| = 0,$$

a result which readily follows by an application of the substitution formula; S_1 becomes simply $1 \dot{-} Sx = 0$, and for S_2, S_3 we have the equations

$$(1 \dot{-} |Sx,Sy|)|x,y| = 0, \quad (1 \dot{-} |x,y|)|Sx,Sy| = 0$$

both of which are consequences of the proved equation $|Sx,Sy| = |x,y|$. The induction schema I is translated by the schema

$$\frac{F(0) = 0, \quad (1 \dot{-} F(n))\, F(Sn) = 0}{F(n) = 0}$$

to which we have already drawn attention.

As is to be expected the deduction theorem is valid also for the equation calculus, and it can be shown that if the equation $P = Q$ can be proved, treating a hypothetical equation $F = G$ as if it were a proved equation, then, provided that the derivation of $G = 0$ does not apply the schema Sb_1 to variables in F, the equation

$$(F = G) \rightarrow (P = Q)$$

is provable without the aid of the hypothesis $F = G$.

A similar result holds for more than one hypothesis; for instance if $P = Q$ is derivable from $F_1 = G_1$ and $F_2 = G_2$ with the same restriction on the use of Sb_1 for the variables in F_1, F_2, G_1, G_2, then

$$[(F_1 = G_1) \ \& \ (F_2 = G_2)] \rightarrow (P = Q)$$

is provable without hypothesis. The need for the restriction on the use of Sb_1 is obvious enough, for without this restriction, from the equation

$$x + 1 = 2$$

we derive

$$0 + 1 = 2$$

but

$$(x + 1 = 2) \rightarrow (1 = 2)$$

is certainly not provable since $(1 + 1 = 2) \rightarrow (1 = 2)$ is patently false.

Counting is formalizable in recursive arithmetic by means of a counting operator N_x^n. We interpret $N_x^n Px$ as the number of true statements amongst $P0, P1, \ldots, Pn$, and define

$$N_x^0 Px = 1 \dot{-} p(0)$$
$$N_x^{n+1} Px = N_x^n Px + \{1 \dot{-} p(n+1)\}$$

where $p(n)$ is the representing function of the recursive predicate Px.

To show that the operator N_x^n counts correctly we sketch the proof of the fundamental theorem

$$N_x^n(x \leqq n) = n + 1.$$

We start by establishing the implication

$$(x > n) \rightarrow \{N_y^n(y = x) = 0\}$$

which we abbreviate as Pn.

$P0$ is $$(x > 0) \rightarrow \{1 \dot{-} |x,0|\}$$

i.e. $$\{1 \dot{-} (1 \dot{-} x)\}(1 \dot{-} x) = 0$$

which follows from the proved equation $(1 \dot{-} x)x = 0$. From the implication

$$(x > n + 1) \to (x > n)$$

we readily derive

$$\{Pn \;\&\; x > n + 1\} \to \{N_y^n(y = x) = 0\}$$

and thence, since $x > n + 1$ implies $1 \dot{-} |x,n + 1| = 0$ we can prove

$$\{Pn \;\&\; x > n + 1\} \to \{N_y^{n+1}(y = x) = 0\}$$

and thence

$$Pn \to P(n + 1)$$

which completes a proof of Pn by induction.

We turn next to the equation

$$N_x^p(x \leqq n + 1) = N_x^p(x \leqq n) + N_x^p(x = n + 1);$$

the proof of this for $p = 0$ is obvious; to complete the proof we use the equation

$$(p \dot{+} n) = \{1 \dot{-} (Sp \dot{-} n)\} + \{1 \dot{-} |p,n|\}$$

which we proved above. We have

$$\begin{aligned} N_x^{p-1}(x \leqq n + 1) &= N_x^p(x \leqq n + 1) + \{1 \dot{-} (p \dot{-} n)\} \\ N_x^{p+1}(x \leqq n) &+ N_x^{p+1}(x = n + 1) \\ &= N_x^p(x \leqq n) + N_x^p(x = n \dot{-} 1) \\ &\quad + \{1 \dot{-} (Sp \dot{-} n)\} + \{1 \dot{-} |p,n|\} \\ &= N_x^p(x \leqq n) + N_x^p(x = n + 1) + \{1 \dot{-} (p \dot{-} n)\} \end{aligned}$$

and the proof is completed by the equalizing rule.

Finally we consider the theorem

$$N_x^n(x \leqq n) = n + 1.$$

For $n = 0$, we have

$$N_x^0(x \leqq 0) = 1 \dot{-} (0 \dot{-} 0) = 1 = 0 + 1;$$

furthermore

$$\begin{aligned} N_x^{n+1}(x \leqq n + 1) &= N_x^{n+1}(x \leqq n) + N_x^{n+1}(x = n + 1) \\ &= N_x^n(x \leqq n) + \{1 \dot{-} ((n + 1) \dot{-} n)\} + 1 \\ &= N_x^n(x \leqq n) + 1 \end{aligned}$$

whence

$$N_x^n(x \leqq n) = N_x^0(x \leqq 0) + n = n + 1.$$

Arithmetical relations

We return now to a study of the full number system Z. We have already remarked that Z is deficient in function signs but that this deficiency is not as serious as it may at first seem. This is a consequence of the following theorem.

If $f(x)$ is a primitive recursive function (of one or more variables) *then the relation $y = f(x)$ is arithmetical*, that is to say, there is a formula in the system Z which expresses this relation.

We notice first that the result holds for each of the initial functions Zx, Ix, and Sx, for the relations $y = Zx$, $y = Ix$, and $y = Sx$ are represented in Z by the formulae $y = 0$, $y = x$, and $y = Sx$.

If $f(x)$ is obtained by substitution, say

$$f(x) = h(u(x),v(x))$$

where the relations $y = h(u,v)$, $y = u(x)$ and $y = v(x)$ are arithmetical, then the relation $y = f(x)$ may be expressed as

$$(\exists u)(\exists v)(\mathrm{U}(x,u) \,\&\, \mathrm{V}(x,v) \,\&\, \mathrm{H}(u,v,y))$$

where $\mathrm{U}(x,u)$, $\mathrm{V}(x,v)$ and $\mathrm{H}(u,v,y)$ stand for the representations in Z of the relations $u = u(x)$, $v = v(x)$ and $y = h(u,v)$ respectively. There remains to consider the case when $f(x,t)$ is primitive recursive in recursive functions $a(t)$, $b(x,t,u)$. We have shown that, given any sequence of numbers $v_0, v_1, v_2, \ldots, v_p$ we can find numbers x and d so that

$$\rho(x, 1 + (i + 1)d) = v_i$$

for each value of i from 0 to p. Let us for brevity write $\gamma(x,d,i)$ for $\rho(x, 1 + (i + 1)d)$ and let $\Gamma(x,d,i,y)$ stand for the representation in Z of the relation

$$\rho(x, 1 + (i + 1)d) = y$$

which we already know to be arithmetical.

The introductory equations of $f(x,t)$ are

$$f(0,t) = a(t),\ f(Sx,t) = b(x,t,f(x,t)).$$

Let $v_0, v_1, v_2, \ldots, v_p$ be the values of $f(0,t), f(1,t), \ldots f(p,t)$ for given values of $p > 0$ and t, so that

$$v_0 = a(t),\ v_1 = b(0,t,v_0),\ v_2 = b(1,t,v_1), \ldots, v_p = b(p - 1,t,v_{p-1}).$$

As we have already observed, there are numbers c,d so that

$$\gamma(c,d,i) = v_i$$

for all values of i from 0 to p, that is to say,

$$\gamma(c,d,0) = a(t) \;\&\; (\forall i)\{i < p \to \gamma(c,d,i+1) = b(i,t,\gamma(c,d,i))\}.$$

Accordingly the relation $y = f(p,t)$ implies that

$$\begin{aligned}(\exists c)(\exists d)(\exists a)[\gamma(c,d,p) &= y \\ &\;\&\; \gamma(c,d,0) = a \;\&\; a = a(t) \\ &\;\&\; (\forall i)\{i < p \to (\exists u)(\exists v)(\gamma(c,d,i) = u \\ &\;\&\; \gamma(c,d,i+1) = v \;\&\; b(i,t,u) = v)\}],\end{aligned}$$

and the converse is obviously true.

This informal argument may be paralleled by a formal proof in Z of the formule F:

$$\begin{aligned}(\exists c)(\exists d)(\exists a)[\Gamma(c,d,\mathbf{p},\mathbf{y}) &\;\&\; \Gamma(c,d,0,a) \;\&\; a = a(\mathbf{t}) \\ &\;\&\; (\forall i)\{i < \mathbf{p} \to (\exists u)(\exists v)(\Gamma(c,d,i,u) \\ &\;\&\; \Gamma(c,d,i+1,v) \;\&\; \mathrm{B}(i,\mathbf{t},u,v))\}],\end{aligned}$$

where $\mathbf{p}$, $\mathbf{t}$, $\mathbf{y}$ stand for the representations in Z of any integers p, t, y satisfying $y = f(p,t)$, and $\mathrm{B}(i,t,u,v,)$ stands for the representation in Z of the arithmetical relation $b(i,t,u) = v$. Formula F is therefore the representative in Z of the relation $y = f(p,t)$.

It follows of course that if $\mathrm{R}(x,y,\ldots)$ is a primitive recursive relation then it is arithmetical. For there is a primitive recursive function $r(x,y,\ldots)$ such that R is equivalent to

$$r(x,y,\ldots) = 0;$$

hence if $\mathscr{R}(x,y,\ldots,w)$ represents the relation $w = r(x,y,\ldots)$ in Z then $\mathscr{R}(x,y,\ldots,0)$ represents $\mathrm{R}(x,y,\ldots)$ in Z.

CHAPTER IV

THE INCOMPLETENESS OF ARITHMETIC

Gödel Numbering, and the Arithmetization of Syntax. Undecidable Statements. Impossibility of Characterizing the Natural Numbers by an Axiomatic System. The Decision Problem. The Undecidability of Arithmetic and the Undecidability of Predicate Logic.

The incompleteness of Arithmetic

We shall apply the result obtained in the last chapter to prove that the number system Z is incomplete. First, however, we shall show how the syntax of Z may be expressed in Z by means of an enumeration of the elements, formulae, and proofs of Z.

The symbols, formulae, and proofs of the formal system Z may be numbered off in the following way. The signs

$$0 \quad S \quad = \quad \& \quad \vee \quad \rightarrow \quad \rightarrow \quad \forall \quad \exists \quad + \quad \cdot \quad (\quad)$$

are assigned the odd numbers from 3 to 27 respectively. Sentence variables are given the numbers $4n + 29$, and number variables the numbers $4n + 31$ for values $0,1,2, \ldots$ of n.

To number the formulae of Z we first number the symbols of the formula and then correlate the number of the formula with the sequence of the numbers of the symbols, uniquely, as follows: if $k_1, k_2, \ldots, k_n$ are the numbers of the symbols of a formula F then the number assigned to F is

$$2^{k_1} \cdot 3^{k_2} \cdot 5^{k_3} \cdot 7^{k_4} \cdot \ldots p_n^{k_n}$$

where p_n is the nth prime number. Since factorization is unique (when the prime factors are held in increasing order of magnitude) the signs of which a formula is composed may be recovered from the number of the formula. All individual symbols have odd numbers and sequences of symbols have even numbers. The numbering process in fact constitutes a code in which each message has a unique number, the 'letters' of the message being discovered simply by factorizing the message number. The length of the message of number x, that is the number of

letters in the message, is the least $y \leqq x$ such that $\{x\}_y = 0$. We denote this function by $L(x)$, so that $L(x)$ is a primitive recursive function.

The axioms of Z are the four axioms of the sentence calculus, the two axioms of the predicate calculus and the eight axioms for equality and the functions S, +, and . ; these axioms are expressed by formulae whose numbers may readily be determined. We are not in fact interested in the actual numbers and may denote the numbers of the fourteen axioms simply by A_1, A_2, . . ., A_{14}, so that the recursive predicate 'x is the number of an axiom', denoted by $Ax(x)$ is just the equation

$$|x,A_1| \cdot |x,A_2| \cdot \ldots |x,A_{14}| = 0.$$

To handle the inference rules we introduce the auxiliary operation $x \wedge y$ defined so that if

$$x = 2^{k_1} \cdot 3^{k_2} \cdot \ldots \cdot p_r^{k_r}$$

and

$$y = 2^{l_1} \cdot 3^{l_2} \cdot \ldots \cdot p_s^{l_s}$$

then

$$x \wedge y = 2^{k_1} \cdot 3^{k_2} \cdot \ldots \cdot p_r^{k_r} \cdot p_{r+1}^{l_1} \cdot \ldots \cdot p_{r+s}^{l_s};$$

$x \wedge y$ is the number of the sentence formed by writing the sentence of number y after the sentence of number x. It is readily verified that $x \wedge y$ is a primitive recursive function.

It follows that $B(x) = 2^{25} \wedge x \wedge 2^{27}$ is the number of the sentence obtained by enclosing sentence numbered x within brackets, and $Neg(x) = 2^{13} \wedge B(x)$ is the number of $\neg(A)$ where x is the number of A.

If x and y are the numbers of statements p and q then the number of the implication $(p) \rightarrow (q)$ is $Imp(x,y) = B(x) \wedge 2^{15} \wedge B(y)$; and if τ is the number of the variable t then the number of the universal sentence

$$(\forall t)(p)$$

is

$$U(\tau,x) = B(2^{17} \wedge \tau) \wedge B(x)$$

with a similar recursive formula for the existential statement $(\exists t)(p)$.

Accordingly the equation

$$x = Imp(y,z),$$

denoted by $Mp(x,y,z)$, says that z is the number of a statement which is an inference by *modus ponens* from the statements with

numbers x and y respectively. In a similar way we can formulate equations to express inference by the rules

$$\frac{S \to A(t)}{S \to (\forall t)A(t)'} \qquad \frac{A(t) \to S}{(\exists t)A(t) \to S};$$

we shall represent these equations by $\mathrm{Un}(x,y)$, $\mathrm{Ex}(x,y)$ which say that y is the number of the sentence obtained from the sentence numbered x by binding a variable, previously free, by a universal or an existential operator as the case may be. So, too, inference by induction may be represented by a formula $\mathrm{Ind}(x,y)$.

Thus the disjunction

$$\mathrm{Mp}(x,y,z) \vee \mathrm{Un}(x,z) \vee \mathrm{Ex}(x,z) \vee \mathrm{Ind}(x,z)$$

says that sentence numbered z follows from sentence numbered x or from sentences numbered x and y. We denote this disjunction by

$$\mathrm{I}(x,y,z).$$

We are now ready to formulate the predicate 'x is the number of a proof'. This predicate, denoted by $\mathrm{Pf}(x)$, is definable as

$$\mathrm{L}(x) > 0 \;\&\; (\forall n)[0 < n \leqq \mathrm{L}(x) \to$$
$$(\mathrm{Ax}\{x\}_n \vee (\exists p)(\exists q)(0 < p < n \;\&\; 0 < q < n \;\&\; \mathrm{I}(\{x\}_p,\{x\}_q,\{x\}_n)))]$$

which says that in the prime factorization of the number x each exponent n is either the number of an axiom or is preceded by exponents p and q which are numbers of sentences from which sentence number n follows by a rule of inference.

Finally we define 'x is the number of a proof of sentence numbered y', denoted by $\mathrm{Pf}(x,y)$, as

$$\mathrm{Pf}(x) \;\&\; y = \{x\}_{\mathrm{L}(x)}.$$

We have shown that $\mathrm{Pf}(x,y)$ is a primitive recursive relation. The next task is to show that the effect of *substituting for a variable* may also be expressed by a primitive recursive function.

The predicate 'n is the number of a variable in Z', is equivalent to 'n is of the form $4r \div 31$' and so may be expressed as

$$(\exists r)(r < n \;\&\; n = 4r + 31)$$

which is primitive recursive. We shall denote the representing function of this predicate by $v(n)$. And since a term in Z is

either 0 or a variable or is one of the forms Sr, $r + s$, $r \cdot s$, where r and s are terms, the representing function $t(n)$ of the predicate which says that n is the number of a term in Z, clearly satisfies the equation

$$t(n) = |n,3| \cdot v(n) \cdot (t\{n\}_1 + |n,2^5 \cdot 3^{\{n\}_1}|) \cdot (t\{n\}_0 + t\{n\}_2 + |n,2^{\{n\}_0} \cdot 3^{21} \cdot 5^{\{n\}_2}| \cdot |n,2^{\{n\}_0} \cdot 3^{23} \cdot 5^{\{n\}_2}|).$$

Since $\{n\}_0$, $\{n\}_1$ and $\{n\}_2$ are all less than n (for $n > 0$) this constitutes a course-of-values definition of $t(n)$ and therefore $t(n)$ is primitive recursive.

In a similar way, using the fact that every statement of Z takes one of the forms

$$a = b,\ \mathrm{A} \to \mathrm{B},\ \mathrm{A}\ \&\ \mathrm{B},\ \mathrm{A} \vee \mathrm{B},\ \neg \mathrm{A},\ \forall x \mathrm{A}(x),\ \exists x \mathrm{A}(x)$$

where a,b are terms and $\mathrm{A},\mathrm{B},\mathrm{A}(x)$ are statements we may define the primitive recursive relation $\mathrm{St}(n)$ which says that n is the number of a statement in Z.

The same analysis leads to the following characterization of the function $\mathrm{Sb}(e,t,x)$ which gives the number of the formula which is obtained by substituting term number t for variable number x in expression (term or statement) number e.

$$\mathrm{Sb}(x,t,x) = t$$
$$(e = 3) \vee (v(e) = 0\ \&\ e \neq x) \to \mathrm{Sb}(e,t,x) = e$$
$$e = 2^{13} \cdot 3^n \to \mathrm{Sb}(e,t,x) = 2^{13} \cdot 3^{\mathrm{Sb}(n,t,x)}$$
$$e = 2^m \cdot 3^\delta \cdot 5^n \to \mathrm{Sb}(e,t,x) = 2^{\mathrm{Sb}(m,t,x)} \cdot 3^\delta \cdot 5^{\mathrm{Sb}(n,t,x)}$$

(where δ has any one of the values 9,11,15)

$$e = 2^\varepsilon \cdot 3^x \cdot 5^n \to \mathrm{Sb}(e,t,x) = e$$

and

$$y \neq x\ \&\ e = 2^\varepsilon \cdot 3^y \cdot 5^n \to \mathrm{Sb}(e,t,x) = 2^\varepsilon \cdot 3^y \cdot 5^{\mathrm{Sb}(n,t,x)}$$

(where ε has either of the values 17 or 19).

If e is not the number of a term or of a statement we simply define

$$\mathrm{Sb}(e,t,x) = e.$$

It may be shown that these conditions lead to a definition of $\mathrm{Sb}(e,t,x)$ by a course of values recursion, so that $\mathrm{Sb}(e,t,x)$ is primitive recursive.

The numbers of the numerals of Z, namely 0, So, SSo, SSSo, and so on, are respectively 3, $2^5 \cdot 3^3$, $2^5 \cdot 3^5 \cdot 5^3$, $2^5 \cdot 3^5 \cdot 5^5 \cdot 7^3$, and so on. Denoting the nth (odd) prime by p_n, ($p_0 = 2$), and the number of the numeral SS . . . So, with n S's by Z_n then $Z_0 = 3$, $Z_1 = 2^5 \cdot 3^3$ and for $n \geqq 1$, $Z_{n+1} = p_n^2 p_{n+1}^3 Z_n$, so that Z_n is primitive recursive, and therefore

$$\mathrm{St}(e,n,x) = \mathrm{Sb}(e,Z_n,x)$$

is primitive recursive. $\mathrm{St}(e,n,x)$ is the number of the formula which results by substituting the numeral SS . . . So with n S's for variable number x in formula number e. We denote the numeral SS . . . So with an assigned number k of S's for short by z_k.

Since the relation $\mathrm{Pf}(x,y)$ and the function $\mathrm{St}(e,n,x)$ are primitive recursive, therefore the relation

$$\mathrm{Pf}(x,\mathrm{St}(y,y,z))$$

is primitive recursive and so expressible in Z by means of a formula $\mathscr{P}(x,y,z)$ say.

Let the number of the variable y be η and let $\mathscr{G}(y)$ denote the formula

$$(\forall x) \neg \mathscr{P}(x,y,z_\eta);$$

further let the number of this formula be p.

Since $\mathscr{G}(z_p)$ is the formula which results when variable number η (i.e. y), in formula number p, is replaced by term number Z_p, therefore the number of $\mathscr{G}(z_p)$ is $Sl(p,p,\eta)$. Hence if $\mathscr{G}(z_p)$ is provable in Z, and if k is the number of the proof, then

$$\mathrm{Pf}(k,\mathrm{St}(p,p,\eta))$$

holds, and so

$$\mathscr{P}(z_k,z_p,z_\eta)$$

is provable in Z. It follows that

$$(\exists x)\mathscr{P}(x,z_p,z_\eta)$$

is provable in Z, and therefore $\neg \mathscr{G}(z_p)$ is provable in Z. Thus we have shown that *if* $\mathscr{G}(z_p)$ *is provable in* Z *then* Z *contains a contradiction*. Accordingly, if Z is free from contradiction, $\mathscr{G}(z_p)$ cannot be proved in Z. But the number of the sentence

$\mathscr{G}(z_p)$ is, as we have seen, $\mathrm{St}(p,p,\eta)$ and so we have shown that (if Z is consistent)

$$\neg\, \mathrm{Pf}(k,\mathrm{St}(p,p,\eta))$$

holds for every k; hence

$$\neg\, \mathscr{P}(z_k,z_p,z_\eta)$$

is provable in Z for every z_k.

Let us say that a system is omega consistent (after Cantor's first transfinite ordinal) if, for every predicate $\mathrm{P}(n)$ such that $\neg\, \mathrm{P}(0)$, $\neg\, \mathrm{P}(1)$, $\neg\, \mathrm{P}(2)$, . . . are all provable, the statement $(\exists n)\mathrm{P}(n)$ is *not* provable.

The property of omega consistency is stronger than ordinary consistency; to show this, let L be an omega consistent system, S any statement in L and n a variable not contained in S. Define $\mathrm{T}(n)$ to be equivalent to S for any n; then if $\neg$ S and S are both provable in L (so that L is inconsistent) it follows that $\neg\, \mathrm{T}(0)$, $\neg\, \mathrm{T}(1)$, $\neg\, \mathrm{T}(2)$, . . . are all provable, and since $\mathrm{T}(0)$ is provable, so is $\exists n\mathrm{T}(n)$, which makes L omega inconsistent.

Suppose now that Z is omega consistent. Then, since $\neg\, \mathscr{P}(z_k,z_p,z_\eta)$ is provable in Z for all z_k, $(\exists x)\mathscr{P}(x,z_p,z_\eta)$ is *not* provable, i.e. $\neg\, \mathscr{G}(z_p)$ is not provable.

To sum up, we have shown that (if Z is omega consistent) all the statements

$$\neg\, \mathscr{P}(0,z_p,z_\eta),\quad \neg\, \mathscr{P}(\mathrm{So},z_p,z_\eta),\quad \neg\, \mathscr{P}(\mathrm{SSo},z_p,z_\eta),\quad \ldots$$

are provable in Z, but

$$(\forall x)\neg\, \mathscr{P}(x,z_p,z_\eta)$$

is not provable in Z, nor is its negation $(\exists x)\mathscr{P}(x,z_p,z_\eta)$.

The consistency of Z is expressed by the statement

$$\neg\, (\mathrm{Pf}(x,z)\ \&\ \mathrm{Pf}(y,\mathrm{Neg}\ z))$$

which says that it is false that x is the number of a proof in Z of sentence number z and y is the number of a proof of the contrary of sentence number z. If $\mathscr{B}(x,y)$ and $\mathscr{C}(x,y)$ are the formulae which express the recursive predicates $\mathrm{Pf}(x,y)$ and $\mathrm{Pf}(x,\mathrm{Neg}\,y)$ in the formalism Z,

then

$$\neg\, (\mathscr{B}(z_a,z_c)\ \&\ \mathscr{C}(z_b,z_c))$$

is provable in Z if and only if a,b,c are numerals for which

$$\neg\, (\mathrm{Pf}(a,c)\ \&\ \mathrm{Pf}(b,\mathrm{Neg}\ c))$$

holds. Thus to say that Z is free from contradiction is to say that

$$\rightarrow (\mathscr{B}(z_a,z_c) \ \& \ \mathscr{C}(z_b,z_c))$$

is provable in Z for all z_a, z_b and z_c. However, the proof we have given that if Z is consistent then $\mathscr{G}(z_p)$ is not provable in Z can be paralleled by a formal derivation in Z of $\mathscr{G}(z_p)$ from

$$\rightarrow (\mathscr{B}(x,z) \ \& \ \mathscr{C}(y,z)).$$

It follows that, if Z *is consistent, the proposition which expresses this fact in* Z, *namely,*

$$\rightarrow (\mathscr{B}(x,z) \ \& \ \mathscr{C}(y,z))$$

is not provable in Z.

Of course the possibility remains that every instance of this unprovable statement is provable in Z, that is to say

$$\rightarrow (\mathscr{B}(z_a,z_c) \ \& \ \mathscr{C}(z_b,z_c))$$

may be provable in Z for all z_a,z_b,z_c, and this has in fact been shown to be the case by using *transfinite induction*, which is of course beyond the proof resources of system Z.

The fact that there are formulae $\mathscr{F}(n)$ in Z such that $\mathscr{F}(0)$, $\mathscr{F}(S0)$, $\mathscr{F}(SS0)$, . . ., are all provable in Z whilst $\mathscr{F}(n)$ is not so provable (the predicate $\mathscr{P}(n,z_p,z_n)$ above being an example of such a predicate) shows that the derivation processes in Z, in particular induction and the substitution schema, fail to ensure that the natural numbers are the *only* values which the variables in Z may take.

We can in fact show directly that the natural numbers do not constitute the largest class of objects which satisfy the axioms of number system Z.

To this end we prove that, given any sequence of functions f_1t, f_2t, f_3t, and so on, (recursive or not) there is a monotonic increasing function $g(t)$ and a function $v(i,j)$ so that, for all i,j, one of the relations

$$f_ig(t) < f_jg(t), \quad f_ig(t) = f_jg(t), \quad f_ig(t) > f_jg(t)$$

holds for all $t > v(i,j)$.

We start by arranging all pairs of functions f_r, f_s, with $r < s$, in a simple sequence, starting with the pair f_1, f_2, followed in turn by f_1, f_3; f_1, f_4; f_2, f_3 and so on, the pair f_i, f_j coming

$(e^2 + i)$th or $(e^2 - e + i)$th in the sequence according as $i + j$ is even or odd, where $2e$ is the greatest even number below $i + j$.

For any two functions $\phi(t)$, $\psi(t)$ we can find a monotonic increasing function $v(t)$ such that *one* of the relations

$$\phi v(t) < \psi v(t), \quad \phi v(t) = \psi v(t), \quad \phi v(t) > \psi v(t)$$

holds for all values of t; to determine v we divide the natural numbers into 3 classes $C_<$, $C_=$ and $C_>$, a number t going into the class $C_<$, $C_=$ or $C_>$ according as the relation $<$, $=$ or $>$ holds between ϕt and ψt.

At least one of the three classes is infinite and we take the members of that class in increasing order of magnitude as the values of $v(t)$ for the values 1, 2, 3, . . . of t, and we write V for the relation ($<$, $=$ or $>$) which holds between $\phi v(t)$ and $\psi v(t)$ for all values of t.

Next we define in turn $v_1(t)$, $v_2(t)$, . . . with associated relations V_1, V_2, . . . as follows; $v_1(t)$ and V_1 are the function and relation determined as above so that

$$f_1 v_1(t) \; V_1 f_2 v_1(t)$$

for all values of t; then we take $f_1 v_1$ and $f_3 v_1$ for ϕ and ψ and determine v_2, V_2 so that

$$f_1 v_1 v_2(t) \; V_2 f_3 v_1 v_2(t)$$

for all t. In this way we determine step by step

$$v_3, V_3; \; v_4, V_4; \; \ldots; \; v_n, V_n$$

so that (E) $\quad f_i v_1 v_2 \ldots v_n(t) \; V_n f_j v_1 v_2 \ldots v_n(t)$

for all t, where n is the ordinal of the pair f_i, f_j in the ordering prescribed above. Write

$$g(n) = v_1 v_2 \ldots v_{n-1} v_n(n)$$

so that (taking $t = n$ in formula E)

$$f_i g(n) \; V_n f_j g(n).$$

Similarly substituting $v_{n+1} v_{n+2} \ldots v_{n+p}(t)$ for t in formula E we find that $f_i v_1 v_2 \ldots v_{n+p}(t) \; V_n f_j v_1 v_2 \ldots v_{n+p}(t)$ for all t, and therefore, taking $t = n + p$,

$$f_i g(n + p) \; V_n f_j g(n + p),$$

which proves that

$$f_i g(t) \ \mathrm{V}_n f_j g(t)$$

holds for all $t \geqq n$.

Since $v_{n+1}(t)$ is monotonic increasing (and its values are all different whole numbers) therefore $v_{n+1}(n+1) > n$ so that

$$g(n+1) = v_1 v_2 \ldots v_n v_{n+1}(n+1) > v_1 v_2 \ldots v_n(n) = g(n)$$

which proves that $g(n)$ is monotonic increasing.

By means of this theorem we may introduce a linear ordering between the functions $f_1, f_2, f_3, \ldots$. We write

$$f_i < f_j \quad \text{or} \quad f_i = f_j \quad \text{or} \quad f_i > f_j$$

according as V_n has the value $<$, $=$ or $>$, i.e. according as

$$f_i g(t) < f_j g(t), \quad f_i g(t) = f_j g(t) \quad \text{or} \quad f_i g(t) > f_j g(t)$$

for all $t > n$. It is readily verified that

$$f_i = f_j \quad \text{and} \quad f_j = f_k \quad \text{implies} \quad f_i = f_k$$

and $f_i < f_j, f_j < f_k$ implies $f_i < f_k$, as our use of the notation anticipated.

Amongst the functions $f_1, f_2, f_3, \ldots$, we may if we please include all the constant functions $f(t) = 0, f(t) = 1$, and so on, which we may denote simply by 0, 1, and so on, and the identity function $f(t) = t$. Inequalities between natural numbers become the same inequalities between constant functions since for instance if $m < n$ then, regarding m and n as constant functions of t, $m < n$ for *all* t.

Each of the functions $f(t)$ of course satisfies the inequality $0 \leqq f(t)$, and each function $f(t)$ has a *unique* successor $f(t) + 1$, for if there were a function h such that $f < h < f + 1$, then we should have

$$fg(t) < hg(t) < fg(t) + 1$$

for all sufficiently great t, which is of course impossible. Thus the terms of the sequence $f_1, f_2, f_3, \ldots$ shares with the natural numbers the property of having a unique successor but the sequence has members greater than every natural number, since for instance the identity function $f(t) = t$ exceeds every constant function, because $fg(t)$ has the value $g(t)$ which grows with t beyond any assigned amount.

Let us now take for $f_1, f_2, f_3, \ldots$ the sequence of all one variable primitive recursive functions. We shall show that all the true statements of recursive arithmetic remain valid if we take $f_1, f_2, f_3, \ldots$ in place of the natural numbers. Given any recursive function

$$F(x_1, x_2, \ldots, x_n)$$

of natural numbers, we introduce function variables $\phi_1(t)$, $\phi_2(t)$, . . . whose 'values' are the functions $f_1, f_2, \ldots$. For each assignment of values of the variables

$$F(\phi_1, \phi_2, \ldots, \phi_n)$$

becomes a one variable function

$$F(f_{k_1}(t), f_{k_2}(t), \ldots, f_{k_n}(t))$$

say, which we take to be the value of the function

$$F(\phi_1, \ldots, \phi_n)$$

for the values $f_{k_1}, f_{k_2}, \ldots, f_{k_n}$ of the arguments $\phi_1, \phi_2, \ldots, \phi_n$. In particular when we give $\phi_1, \ldots, \phi_n$ for values n *constant functions* $a_1, a_2, \ldots, a_n$, F takes as value the constant function

$$F(a_1, a_2, \ldots, a_n)$$

which is of course exactly the same as the value of $F(x_1, x_2, \ldots, x_n)$ for the values $a_1, a_2, \ldots, a_n$ of the arguments $x_1, x_2, \ldots, x_n$. Thus a function of natural number variables may be interpreted as a function in the space of the functions $f_1, f_2, \ldots$ without changing its significance.

We show next that under this reinterpretation a true equation

$$F(x_1, \ldots, x_n) = 0$$

becomes a true equation in the function space. For if

$$F(x_1, x_2, \ldots, x_n) = 0 \text{ for all } x_1, x_2, \ldots, x_n$$

then

$$F(f_{k_1}(t), f_{k_2}(t), \ldots, f_{k_n}(t)) = 0$$

for all values of t and all suffixes $k_1, k_2, \ldots, k_n$, so that $F(\phi_1, \phi_2, \ldots, \phi_n)$ is equal to the zero function for all values of $\phi_1, \ldots, \phi_n$.

Since every statement in recursive arithmetic may be given the form $F(x_1, \ldots, x_n) = 0$, we have shown that the true statements of recursive arithmetic (and in particular the

axioms) are true not only for natural number arguments but also for argument values chosen from a richer function space.

Thus it becomes readily intelligible that a predicate $\mathscr{P}(x)$ may be such that $\mathscr{P}(o)$, $\mathscr{P}(So)$, $\mathscr{P}(SSo)$, . . . are all provable even when $\mathscr{P}(x)$ is not provable, since recursive arithmetic admits an interpretation in which o, So, SSo, . . . are not *all* the values of x.

By making certain modifications in the setup of the system we may show similarly that the number system Z likewise admits of an interpretation in which o, So, SSo, . . . do not constitute the totality of numerals.

The decision problem

We have seen that sentence logic and certain classes of statements in predicate logic admit a decision procedure, which may be described as a purely mechanical test for validity. In this section we shall look more closely at the nature of a decision procedure with a view to showing that there can be no mechanical test for validity for arithmetic. This is of course hardly surprising since, as is well known, there are many famous unsolved problems in arithmetic, like Goldbach's hypothesis that every even number is a sum of two primes or Fermat's assertion that the equation $x^n + y^n = z^n$ has no solution in positive (non-zero) integers if n exceeds 2.

We shall start from the working assumption that the ordinal recursive functions we considered earlier include every kind of function which could be considered computable, that is to say every function whose value for an assigned argument can be determined in a finite number of steps. The grounds for believing this are heuristic, not mathematical, since of course we are concerned with the matter of fact that no one has ever discovered a computable function which is not ordinal recursive, but there is the added confirmation that the several seemingly diverse formal definitions of computable functions, like λ-definability, which have been devised have all proved to define a class of functions equivalent with the class of ordinal recursive functions.

It is known that ordinal recursive functions of the type to which we have referred are, like primitive recursive functions, definable in Z (and we could take definability in Z as a very

convenient characterization of computable functions). In fact, as we have already remarked, Z_f shares this property with Z. We recall that a function $f(n)$ is said to be definable in Z_f if there is a predicate $F(x,y)$ in Z_f such that

$$F(z_n,z_N) \text{ or } \neg F(z_n,z_N)$$

is provable in Z_f according as $N = f(n)$ or $N \neq f(n)$. We shall show that, in place of F, we can find a Φ in Z_f with the stronger property that

$$(\forall y)(\Phi(z_n,y) \leftrightarrow y = z_{fn})$$

is provable in Z_f.

We take for $\Phi(x,y)$ the predicate

$$F(x,y) \;\&\; (\forall z)(F(x,z) \to z \geqq y),$$

so that the implications

$$\Phi(x,y) \to (F(x,z) \to z \geqq y)$$

and

$$\Phi(x,y) \to F(x,y)$$

are both provable in Z_f. From the first of these follows

$$F(x,z) \to (\Phi(x,y) \to z \geqq y)$$

and thence, by the second,

$$\Phi(x,z) \to (\Phi(x,y) \to z \geqq y)$$

whence in turn we reach

$$\Phi(x,y) \to (\Phi(x,z) \to z \geqq y)$$

and, interchanging y and z

$$\Phi(x,y) \to (\Phi(x,z) \to y \geqq z);$$

from the last two equations we derive

$$\Phi(x,y) \to (\Phi(x,z) \to y = z)$$

and from this follows

(i) $$\Phi(x,y) \to (\forall z)(\Phi(x,z) \to y = z).$$

Next we prove the statement

(ii) $$\Phi(z_n,z_{f(n)}).$$

By hypothesis $F(z_n,z_{f(n)})$ is provable in Z_f; it remains to establish

$$F(z_n,z) \to z \geqq z_{f(n)}$$

which we consider in the form

$$z < z_{f(n)} \to (\multimap F(z_n,z)).$$

Since n is some assigned integer we may readily prove (by an informal induction over n) that

$$z < z_{f(n)} \to (z = z_0) \vee (z = z_1) \vee \ldots \vee (z = z_{f(n)-1});$$

but $$z = z_r \to (\multimap F(z_n,z_r) \to \multimap F(z_n,z))$$

is provable in Z_f, whence, since $\multimap F(z_n,z_r)$ is provable for $r \neq f(n)$, we derive

$$z = z_r \to (\multimap F(z_n,z_r)),\ r = 0,1,2,\ldots,f(n) - 1,$$

whence it follows that

$$z < z_{f(n)} \to (\multimap F(z_n,z))$$

as required.

From (i) and (ii), substituting z_n for x and $z_{f(n)}$ for y we derive

$$(\forall z)(\Phi(z_n,z) \to z = z_{f(n)}).$$

The converse, viz.

$$(\forall z)(z = z_{f(n)} \to \Phi(z_n,z))$$

follows at once from the substitution formula

$$z = z_{f(n)} \to (\Phi(z_n,z_{f(n)}) \to \Phi(z_n,z))$$

and formula (ii).

After these preliminary considerations we come to the proof that arithmetic (as formalized in system Z) has no decision procedure. The proof of undecidability applies equally to Z_f or to Z. We have seen that the statements of a formal system can be numbered off in a quite simple fashion.

Let T be a formal system and let $P(n)$ be the predicate which says that sentence number n is provable in T; then there is a decision procedure for T if, and only if, for each value of n, we can determine in a finite number of steps whether $P(n)$ holds or not. If $p(n)$ is the representing function of the predicate then

we see that T is decidable if, and only if, for each n, we can determine the value of $p(n)$ in a finite number of steps. In other words T is decidable if $p(n)$ is a recursive function and therefore if $P(n)$ is a recursive predicate. Accordingly T is decidable if and only if there is a predicate $T(x)$ in Z_f such that $T(z_n)$ is provable in Z_f if n is the number of a provable formula in T and $\neg T(z_n)$ is provable in Z_f if n is not the number of a provable formula of T.

We have seen that there is a primitive recursive function which gives the number of the expression in Z_f obtained by substituting for a variable in a given expression. In the same way we may show that there is a primitive recursive function, which we again call $St(n,n,\eta)$ which gives the number of the expression in Z_f obtained by substituting z_n for the variable y (with number η we suppose) in expression number n. St is definable in Z_f by a relation $\mathscr{S}(u,v)$, say, which is such that $\mathscr{S}(z_a,z_b)$ is provable if $b = St(a,a,\eta)$, and $\neg \mathscr{S}(z_a,z_b)$ is provable if $b \neq St(a,a,\eta)$. It follows that we can find an expression $\sigma(u,v)$ in Z_f such that

$$(\forall v)(\sigma(z_a,v) \leftrightarrow v = z_b)$$

is provable in Z_f with $b = St(a,a,\eta)$.

Let us now suppose that Z_f is decidable and let $T(x)$ be an expression in Z_f such that $T(z_n)$ is provable or refutable in Z_f according as n is, or is not, the number of a provable statement in Z_f, and let y be a variable not contained in T.

Further let $D(y)$ denote the formula

$$(\forall x)(\sigma(y,x) \rightarrow \neg T(x))$$

and let n be the number of this formula; then the number of the formula denoted by $D(z_n)$ is $St(n,n,\eta)$, which we shall denote by N. We consider the consequences of the hypothesis that $D(z_n)$ is provable in Z, i.e. that

$$(\forall x)(\sigma(z_n,x) \rightarrow \neg T(x))$$

is provable. We conclude first that

$$\sigma(z_n,z_N) \rightarrow \neg T(z_N)$$

is provable. Since $N = St(n,n,\eta)$ therefore $\sigma(z_n,z_N)$ is provable, from which it follows that $\neg T(z_N)$ is provable. If on the other

hand $D(z_n)$ is *not* provable in Z_f, then since N is its number, $\neg T(z_N)$ is provable in Z_f, and so in either case

$$\neg T(z_N)$$

is provable in Z_f.

From the proved formula

$$(\forall v)(\sigma(z_n,v) \to v = z_N)$$

and the substitution formula

$$v = z_N \to (\neg T(z_N) \to \neg T(v))$$

follows

$$\neg T(z_N) \to (\sigma(z_n,v) \to \neg T(v))$$

whence, since $\neg T(z_N)$ is provable,

$$(\forall x)(\sigma(z_n,x) \to \neg T(x))$$

is provable. Thus sentence number N is provable and therefore

$$T(z_N)$$

is provable in Z_f.

Thus the assumption that Z_f is decidable leads to a contradiction in Z_f and therefore, if Z_f is consistent there is no decision procedure for Z_f.

Not only is Z_f undecidable but the same is true of any consistent extension of Z_f, that is any system formed by adding new axioms to Z_f without introducing a contradiction and in particular Z is undecidable. Another example of an undecidable system is the elementary theory of groups, G, whose axioms are

$$x \wedge (y \wedge z) = (x \wedge y) \wedge z$$
$$(\exists z)(x = y \wedge z)$$
$$(\exists y)(x = y \wedge z)$$

and the axioms of predicate logic; the binary operator $\wedge$ is undefined apart from the conditions imposed by the group axioms. If, however, we add the commutative law

$$x \wedge y = y \wedge x$$

to the axioms we obtain a consistent extension of G which is known to be *decidable*. This extension is of course the theory of Abelian groups. The proof of the undecidability of G depends essentially on the fact that the commutative law is not affirmed in G.

From the undecidability of Z_f we readily deduce the undecidability of predicate logic. A necessary preliminary to the derivation is the elimination of the extra-logical predicate and individual symbols and the operators $+$ and $.$ from Z_f. To this end we write the predicate $E(x,y)$ for the equation $x = y$, the predicate $S_0(x,y)$ for the equation $Sx = y$, and the predicates $S_1(x,y,z)$, $P_1(x,y,z)$ for the equations $x + y = z$ and $x \cdot y = z$ respectively. To eliminate the constant o we use axioms S_1 and S_3^* and write $(\forall z) \rightarrow (x = Sz)$ for the equation $x = o$.

Under these transformations axioms S_1, S_3^* become universally valid, axiom S_2 becomes

$$S_0(x,z) \ \& \ S_0(y,z) \rightarrow E(x,y),$$

axiom E becomes

$$E(x,y) \rightarrow \{E(x,z) \rightarrow E(y,z)\},$$

and the axioms for addition and multiplication, axioms A_1, A_2, and P_1, P_2, are transformed into

$$\{(\forall z) \rightarrow S_0(z,w)\} \rightarrow S_1(x,w,x)$$

$$\{S_0(y,u) \ \& \ S_0(v,w) \ \& \ S_1(x,y,v)\} \rightarrow S_1(x,u,w)$$

and

$$\{(\forall z) \rightarrow S_0(z,w)\} \rightarrow P_1(x,w,w)$$

$$S_0(y,u) \ \& \ S_1(v,x,w) \ \& \ P_1(x,y,v) \rightarrow P_1(x,u,w),$$

respectively.

Let the conjunction of these transformations (of which there are only a finite number) be denoted by F, and consider any sentence $\mathscr{S}_f$ in system Z_f. By the given transformations $\mathscr{S}_f$ becomes a predicate sentence S, say, and $\mathscr{S}_f$ is provable in Z_f, if, and only if, S is derivable from F in predicate logic. But by the Deduction Theorem, S is derivable from F if, and only if

$$F \rightarrow S$$

is provable in predicate logic, regarding E, S_0, S_1 and P_1 as predicate variables which are not used as variables in the derivation of S. It follows that if predicate logic were decidable then $F \rightarrow S$ would be decidable; if $F \rightarrow S$ is universally valid, then, since F is provable in Z_f, $\mathscr{S}_f$ is provable in Z_f, and if $\rightarrow (F \rightarrow S)$, i.e. $F \ \& \rightarrow S$ is satisfiable, then $\mathscr{S}_f$ is *not* provable in Z_f. Thus if predicate logic were decidable then Z_f would be decidable and we know that this is not the case.

CHAPTER V

EXTENDED PREDICATE LOGIC

Class Logic. Stratification. Descriptions Operator. Ordered Pairs. Class of Natural Numbers. The Relativity of Class Concepts.

Extended predicate logic

In recursive arithmetic, and in number system Z we have formalizations of arithmetic in which natural number variables are primitive concepts and there is no analysis of the concept of number in terms of other concepts. This is in no respect a shortcoming in these formalizations; some concept must be taken as the basis on which the structure is built and number serves this purpose admirably, if what we are looking for is a foundation for arithmetic. There are, however, branches of mathematics in which the concept of *class* (and particularly *infinite* class) plays an essential part and for these parts of mathematics (modern analysis, set theory, topology) an extended system of logic is needed.

We have already discussed such a class logic in an informal manner. Unfortunately this simple class logic built on the pattern of an intuitive class logic has proved to be self contradictory. For let w be the class of all classes which are not members of themselves, i.e. let w be the class

$$\hat{x} \multimap (x \varepsilon x)$$

(where '$\hat{x}$ . . . x —' denotes the class of all x such that . . . x — holds, and '$x \varepsilon y$' says x is a member of y); then

$$(x \varepsilon w) \leftrightarrow \multimap (x \varepsilon x)$$

is provable for any x. Taking for x the class w itself we are led to the contradiction

$$(w \varepsilon w) \leftrightarrow \multimap (w \varepsilon w).$$

To avoid this, and other less obvious contradictions, formalizations of class logic have of necessity grown exceedingly complicated; and unfortunately no proof of freedom from contradiction has been discovered for any of them. We shall

here briefly outline one of the least restrictive systems of class logic which has the added advantage of being known to be at least as secure from the danger of concealed contradiction as another much narrower system.

The first step in the formulation is to set the universal and existential quantifiers free to operate upon class variables $\alpha, \beta, \gamma, \ldots$. The resulting extended predicate logic is based upon the axioms of sentence logic and the three axioms

$$(\forall\alpha)(\phi \to \psi) \to ((\forall\alpha)\phi \to (\forall\alpha)\psi);$$
$$\phi \to (\forall\alpha)\phi,$$

where α is a variable which is not free in ϕ;

$$(\forall\alpha)\phi \to \phi',$$

where ϕ' is obtained from ϕ by substituting α' for each free occurrence of α.

The only rule of inference is *modus ponens*

$$\frac{\phi,\ \phi \to \psi}{\psi}.$$

To exclude the contradiction of the class of classes which are not members of themselves we impose a membership condition on objects from which classes may be formed.

An object which satisfies the membership condition is called an element. The condition for an object x to be an element is that there should be a class of which it is a member, so that x is an element if $(\exists y)(x \in y)$ holds. Thus if w is the class of all elements x such that $\neg(x \in x)$, then the meaning of $x \in w$ is not just that $\neg(x \in x)$ but that

$$(\exists y)(x \in y)\ \&\ \neg(x \in x);$$

hence instead of the equivalence

$$x \in w \leftrightarrow \neg(x \in x)$$

with the contradictory consequence

$$w \in w \leftrightarrow \neg(w \in w),$$

we have the equivalence

$$x \in w \leftrightarrow (\exists y)(x \in y)\ \&\ \neg(x \in w)$$

from which follows

$$w \varepsilon w \leftrightarrow (\exists y)(w \varepsilon y) \;\&\; \rightharpoonup (w \varepsilon w)$$

and this is equivalent to

$$\rightharpoonup (\exists y)(w \varepsilon y)$$

so that the contradiction has been transformed into a proof that w is not an element, i.e. is not eligible for class membership.

If we introduce V to denote the class of all elements, i.e.

$$\hat{x}(\forall y)(y \varepsilon x \leftrightarrow y \varepsilon x),$$

where the condition

$$(\forall y)(y \varepsilon x \leftrightarrow y \varepsilon x)$$

is of course satisfied for *all* x, then we can readily show that the membership condition $(\exists y)(x \varepsilon y)$ is equivalent to $x \varepsilon \mathrm{V}$ which may therefore be taken as a convenient normal form for the membership condition.

However, without some axiom to provide us with a supply of elements we should be incapable of developing a logic of classes. Of course to take *all* statements of the form

$$\hat{\alpha}\phi \varepsilon \mathrm{V}$$

as axioms would simply nullify our membership condition, since this would just make every class an element. Some intermediate stage between taking all such expressions as axioms, and taking none of them must be found, and the best that can be done is to impose some test on ϕ (and α) which will leave as many safe statements as possible in the class of axioms whilst excluding any that are actually refutable. Such a test is provided *in part* by a process of 'potential typing' according to which a formula

$$\hat{\alpha}\phi \varepsilon \mathrm{V}$$

is an axiom if ϕ is stratified, that is to say if the variables in ϕ can be graded into different types in such a way that for each term $x \varepsilon y$ in ϕ, y has a type number one higher than that of x. For instance

$$(\forall y)(x \varepsilon y \rightarrow (\forall z)(y \varepsilon z \leftrightarrow x \varepsilon y))$$

is stratified since we may take x in type 1, y in type 2, and z in type 3, but

$$x \varepsilon y \,\&\, y \varepsilon x$$

is not stratified.

Stratification alone is not, however, a sufficient restriction, for $x \varepsilon y$ is stratified and if

$$\hat{x}(x \varepsilon y) \varepsilon \mathrm{V}$$

were an axiom then it can be shown that $(\forall y)(y \varepsilon \mathrm{V})$ follows and so once again the membership condition would be nullified. An additional condition which appears to be sufficient is to restrict the free variables and the bound variables in ϕ to elements. For the bound variables this is accomplished by changing each term $(\forall \alpha)\psi$ within ϕ into $(\forall \alpha)(\alpha \varepsilon \mathrm{V} \to \psi)$ and each term $(\exists \alpha)\psi$ into $(\exists \alpha)(\alpha \varepsilon \mathrm{V} \,\&\, \psi)$. Thus as the axioms supplying elements we take all sentences of the form

$$(\forall \beta_1)(\forall \beta_2) \ldots (\forall \beta_n)(\beta_1 \varepsilon \mathrm{V} \,\&\, \beta_2 \varepsilon \mathrm{V} \,\&\, \ldots \,\&\, \beta_n \varepsilon \mathrm{V} \to \hat{\alpha}\phi \varepsilon \mathrm{V})$$

which reduces to $\hat{\alpha}\phi \varepsilon \mathrm{V}$ when $n = 0$ where α, β_1, . . ., $(\beta_n$ are *all* the free variables in ϕ and ϕ is formed from a stratified formula by restricting all bound variables to elements.

For instance since the formula

$$(\exists z)(x \varepsilon z \,\&\, (\forall y)(z \varepsilon y \to y \varepsilon w))$$

is stratified, with y and z as bound variables and w free, therefore

$$(\forall w)\{w \varepsilon \mathrm{V} \to x(\exists z)$$
$$[z \varepsilon \mathrm{V} \,\&\, x \varepsilon z \,\&\, (\forall y)(y \varepsilon \mathrm{V} \to (z \varepsilon y \to y \varepsilon w))] \varepsilon \mathrm{V}\}$$

is an axiom.

Next we need axioms supplying the class formed by all elements which satisfy a given property, and for these we take all sentences of the form

$$(\exists \beta)(\forall \alpha)(\alpha \varepsilon \beta \leftrightarrow \alpha \varepsilon \mathrm{V} \,\&\, \phi)$$

where β is not free in ϕ.

Finally we formulate a substitution axiom in the form

$$\alpha = \alpha' \to (\phi \to \phi')$$

where ϕ' is formed by substituting α' for α in ϕ, and $\alpha = \alpha'$ is defined as

$$(\forall \gamma)(\gamma \varepsilon \alpha \leftrightarrow \gamma \varepsilon \alpha').$$

The formal development of arithmetic in class logic is a long and tedious process and we shall preface the details of this formal development by an informal discussion.

The number of a class α, Nc α, is defined as the class of all elements similar to α. The individual numbers 0, 1, 2 and so on are defined in turn. 0 may be taken as $\iota\Lambda$, where Λ is the null class, i.e. the class $\hat{x} \prec (x = x)$, and $\iota\alpha$ means the class whose only member is α.

The definitions of 1, 2, 3 and so on are obtained by taking z to be 0, 1, 2, . . . in the following definition: If z is a natural number then $z + 1$ is defined as the number of the class of all those elements which contain a member whose removal leaves a class of number z.

The relation of $z + 1$ to z is called the successor relation S.

Given some relation R we introduce now the relations R^2, R^3, R^4 and so on (known as the relative powers of R); denoting by $a\mathrm{K}b$ the fact that a bears the relation K to b, we define in turn

$$a\mathrm{R}^2b \leftrightarrow (\exists c)(a\mathrm{R}c \,\&\, c\mathrm{R}b)$$

$$a\mathrm{R}^3b \leftrightarrow (\exists c)(\exists d)(a\mathrm{R}c \,\&\, c\mathrm{R}d \,\&\, d\mathrm{R}b)$$

and so on.

For instance if $a\mathrm{R}b$ says that a is the father of b then $a\mathrm{R}^2b$ says that a is the grandfather of b, and $a\mathrm{R}^3b$ that a is the great-grandfather of b.

Then the sum $x + y$ is defined as the number which bears the relation S^y to x, and if A_x denotes the relation S^x then the product $x \times y$ is the number which the relation A_x^y to 0. For instance $x + 3$ bears the relation S^3 to x and so is the successor of the successor of the successor of x; i.e.

$$x + 3 = ((x + 1) + 1) + 1.$$

And $3x$ bears the relation A_x^3 to 0, where A_x is the relation of x to 0, so that if y and z are such that $3x$ bears the relation A_x to y and y the relation A_x to z and z the relation A_x to 0, then we find in turn

$$z = \mathrm{S}^x0 = x,\ y = \mathrm{S}^xx = x + x,$$

and

$$3x = \mathrm{S}^x(x + x) = (x + x) + x.$$

We come now to an outline of the details of the formalization of the foregoing processes.

To start with we define the notation $\hat{\alpha}\phi$ *contextually* in the sentence

$$\beta \varepsilon \hat{\alpha}\phi \leftrightarrow (\exists\gamma)(\beta \varepsilon \gamma \ \& \ (\forall\alpha)(\alpha \varepsilon \gamma \rightarrow \phi))$$

which says in effect that β belongs to $\hat{\alpha}\phi$ if and only if there is a class γ which contains β and contains α only if ϕ is satisfied by α. Next we introduce the *descriptions operator* $(\iota\alpha)\phi$ which is defined to be

$$\hat{\beta}(\exists\gamma)(\beta \varepsilon \gamma \ \& \ (\forall\alpha)(\alpha = \gamma \leftrightarrow \phi))$$

i.e. the unique class γ which has the property ϕ (if there is one) so that $(\iota\alpha)\phi$ is the class satisfying ϕ if there is just one, or is the null class if ϕ is not satisfiable, or is satisfied by more than one class.

The class whose only member is α, denoted by $\iota\alpha$, is defined as $\hat{\beta}(\beta = \alpha)$. Whether α itself is an element or not, $\iota\alpha$ may be shown to be an element.

The class of all members of a class y, except x, denoted by y/x, is defined as

$$\hat{z}(z \varepsilon y \ \& \ \neg(z = x));$$

y/x is an element if and only if y itself is.

A class x is said to be contained in a class y, denoted by $x < y$, if $(\forall\alpha)(\alpha \varepsilon x \rightarrow \alpha \varepsilon y)$.

We come next to the very important concept of an *ordered pair*. An ordered pair $[x, y]$ is defined to be the class whose only members are ιx and $\iota\overline{xy}$ where $\overline{xy}$ is the class whose only members are ιx and ιy. It is necessary to take ιx and $\iota\overline{xy}$ instead of x and $\overline{xy}$ as seems more natural, because we wish to ensure that $[x,y]$ is an element.

By means of ordered pairs we readily define the concept 'x stands in the relation R to y', denoted by $R(x, y)$, as

$$x \varepsilon V \ \& \ y \varepsilon V \ \& \ [x, y] \varepsilon R$$

so that a relation is simply a class of ordered pairs.

Just as the class of all α with a property ϕ is denoted by $\hat{\alpha}\phi$ so the relation of α to β when α,β have the property ϕ is denoted by $\hat{\alpha}\hat{\beta}\phi$ and defined to be

$$\hat{\gamma}(\exists\alpha)(\exists\beta)(\alpha \varepsilon V \ \& \ \beta \varepsilon V \ \& \ \gamma = [\alpha,\beta] \ \& \ \phi).$$

Given a relation y the object which stands in this relation

to x is denoted by $y(x)$, which is of course an abbreviation for $(\iota\alpha)y(\alpha,x)$.

Given a relation y and a class x, the class of all elements $y(\alpha)$ for an α belonging to x is called the *image* of x by the relation y, and denoted by $y((x))$. Thus $y((x))$ stands for

$$\hat{\alpha}(\exists\beta)(\,y(\alpha,\beta)\ \&\ \beta\ \varepsilon\ x).$$

The image of ιz, the class whose sole member is z, by a relation x is simply the class of all elements which have the relation x to z, i.e.

$$x((\iota z)) = \hat{y}(x(\,y,z)).$$

In terms of the image of a class in a relation we define the *ancestral* of a relation x, denoted by x^* to be the relation

$$\hat{z}\hat{w}(\forall y)(x((\,y)) < y\ \&\ w\ \varepsilon\ y\ \&\ z\ \varepsilon\ y).$$

This formulation of the ancestral relation constitutes a logical analysis of the notion 'x^* is the relation of z to w when either z is w, or z bears the relation x to w, or z bears the relation x to something which bears the relation x to w, and so on'.

We are now able to formulate the definition of the successor of a natural number x, denoted by $S(x)$. $S(x)$ is

$$\hat{y}(\exists z)(\,z\ \varepsilon\ y\ \&\ y/z\ \varepsilon\ x).$$

Defining 0 as $\iota\Lambda$, we have in turn $1 = S(0)$, $2 = S(S(0))$ and so on. By means of the ancestral of the successor we define the *class of all natural numbers*, denoted by Nn, to be $S^*((\iota 0))$, so that to be a natural number is to be 0 or to bear the relation S to 0, or the relation S to something which bears the relation S to 0, and so on. This is neatly summed up in the theorem

$$S((Nn)) < Nn$$

which says that the successor of a natural number is a natural number.

Mathematical induction is provable in the form

$$(\forall\alpha)(\phi \to \phi')\ \&\ \phi_0\ \&\ \zeta\ \varepsilon\ Nn \to \psi$$

where ψ, ϕ' and ϕ_0 are obtained from ϕ by substituting ζ, $S(\alpha)$ and 0 respectively for each free occurrence of α in ϕ. (We omit the proof.)

To define the relative powers of a relation x we introduce the relation x_s, which is the relation of the pair $[x(m),n]$ to the pair $[m,S(n)]$ where n is a natural number. The definition of x_s is

$$\hat{\alpha}\hat{\beta}(\exists\alpha')(\exists\beta')(\exists\gamma)(\alpha = [\alpha',\gamma] \\ \& \ \beta = [\beta',S(\gamma)] \ \& \ \gamma \ \varepsilon \ Nn \ \& \ x(\alpha',\beta')).$$

Finally we reach the definition of x^y as

$$\hat{\alpha}\hat{\beta}(x_s^*([\alpha,0],[\beta,y])).$$

To illustrate this definition let a,b,c,d,e be a succession of objects such that, for some relation x,

$$x(a,b), \ x(b,c), \ x(c,d) \text{ and } x(d,e);$$

then x_s is the relation of any pair to its successor in the series

$$[a,0], \ [b,1], \ [c,2], \ [d,3], \ [e,4]$$

and so each of

$$x_s^*([a,0], \ [b,1]), \ x_s^*([a,0], \ [c,2]), \\ x_s^*([a,0], \ [d,3]), \ x_s^*([b,1], \ [e,4]) \text{ etc,}$$

holds in virtue of the definition of the ancestral of the relation x_s. Thus the definition of x^y makes x^2 the relation of a to c, x^3 the relation of a to d (or b to e) and so on, as desired. As we have already seen addition is definable by means of S^y, with $x + y$ standing for $S^y(x)$ and similarly xy may be defined to be $(S^x)^y(0)$.

This is as far as we shall carry the development of arithmetic in class logic.

We conclude by noting a remarkable paradox to which every formalization of class logic is subject. The paradox concerns the class of all subclasses of a class.

We may readily prove that the class of all subclasses of the class of natural numbers Nn is not similar to Nn. For if the class of all subclasses were similar to Nn then each subclass would have a unique number corresponding to it (which we may call the number of the subclass) and each number would be the number of just one subclass. Denote the subclass of number n by σ_n. Consider now subclass Σ determined as follows: Σ contains the natural number n if subclass σ_n does not contain the number n, and Σ does not contain n if σ_n does contain n. Let N

be the number of Σ. Then Σ is the same subclass as σ_N; but if σ_N contains the natural number N then Σ does not, and *vice-versa*, which proves that Σ is *not* the same subclass as σ_N. This contradiction shows that the class of all subclasses of N*n* is *not* similar to N*n*, and the proof we have given is formalizable in the foregoing system of class logic. But as we have already remarked in the proof of the completeness of predicate logic, every statement or sequence of statements which are all jointly satisfiable in some non-empty domain are jointly satisfiable in the domain of the natural numbers. Now the above formalization of class logic is based on a denumerable infinity of axioms (each axiom form constituting a sequence of axioms), and therefore (if they are satisfiable at all) they are satisfiable in the domain of the natural numbers. *Thus we may interpret the primitive notions so that there is only a denumerable infinity of classes, the axioms all remaining true, even though a theorem in the system asserts that the totality of classes is not denumerable.* It appears that we must either give up the attempt to formalize the class concept by means of a finite number or denumerable infinity of axioms or must accept the position that the non-denumerability of a class may be simply an accidental feature of a formal system, a class which is non-denumerable in our axiomatization becoming denumerable in another. The paradox may be regarded as an incompleteness theorem for class theory, showing that no consistent formalization of class theory by a finite number or denumerable infinity of axioms can express the totality of functions, since the function which enumerates the classes of the theory is not expressible in the formalization.

The Frege-Russell definition of number

Frege's major work, the *Grundgesetze der Arithmetik*, was published in two volumes in 1893 and 1903. An earlier work which does not introduce symbolic logic is *Die Grundlagen der Arithmetik* (1884), (English translation by J. L. Austin, 1950). The corresponding works by Bertrand Russell are *Principia Mathematica* (with A. N. Whitehead, 3 volumes, 1910, 1912, and 1913), and *The Principles of Mathematics* (1903). Russell's *Introduction to Mathematical Philosophy* presents a very readable non-technical introductory account of his outlook on the foundation of mathematics.

Axiomatic sentence logic

The earliest accounts on modern lines are by Frege and Russell, anticipated in some respects by C. S. Peirce and E. Schröder. The independence of the axioms was proved by P. Bernays; completeness was proved by J. Łukasiewicz (and others). The axiom system for intuitionistic logic was discovered by A. Heyting in 1930.

Bracket free notation

This logical notation was introduced by J. Łukasiewicz.

Natural inference

This method of presenting logical proof was devised by G. Gentzen and published in 1934 in the *Mathematische Zeitschrift*, Vol. XXXIX, pp. 176–210, 405–431.

Predicate logic

The *independence* of a set of postulates for predicate logic was first established by J. C. C. McKinsey in 1936.

The deduction theorem was discovered by J. Herbrand in 1928. The natural inference method for predicate logic was

[1] Complete bibliographies of works on Mathematical Logic are given in *Abstract Set Theory* by A. A. Fraenkel (Amsterdam, 1953), and in Vol. I of the *Journal of Symbolic Logic* (1936), and in later volumes.

also discovered by G. Gentzen, but we have followed Quine (*Journal of Symbolic Logic*, Vol. XV, 1950, pp. 93–102).

The completeness of the first order predicate calculus was proved by K. Gödel in 1930; the present account is based on the work of L. Henkin. The reduction to the normal form

$$(\exists x_1)(\exists x_2)\ldots(\exists x_k)(\forall y_1)(\forall y_2)\ldots(\forall y_l)\mathscr{A}(x_1,\ldots,x_k,y_1,\ldots,y_l)$$

was discovered by Thoralf Skolem. (For an account of this reduction see *The Principles of Mathematical Logic* by D. Hilbert and W. Ackermann; this is the best introduction to symbolic logic, but it contains no reference to modern natural inference techniques.)

Number theory

The most complete and detailed study of formalized number theory, as brilliantly written as it is comprehensive, is the *Grundlagen der Mathematik* by D. Hilbert and P. Bernays in two volumes, 1934 and 1939.

Recursive arithmetic was introduced by Thoralf Skolem in 1923. The fullest account of the properties of recursive functions is given by Rózsa Péter, *Rekursive Funktionen* (Budapest, 1951). Transfinite recursive functions were introduced by W. Ackermann in 1940. The given normal form for transfinite recursive functions is based on the normal form for general recursive functions obtained by Julia Robinson, and the identification of transfinite recursive functions (of the kind considered here) with general recursive functions definable in Z, established by J. R. Myhill. The calculus of λ-conversion was discovered by Alonzo Church (*The calculi of λ-conversion*, Princeton, 1941). The formalization of recursive arithmetic without logical axioms was discovered by the author and (independently) by H. B. Curry. The present outlined version is new. An account of recursive arithmetic and a system of Mathematical Analysis which may be based on it is given in the author's *Constructive Formalism* (Leicester, 1951).

The proof that the relation $y = f(x)$ is arithmetical if $f(x)$ is *primitive* recursive is due to K. Gödel (based in part on a device introduced by R. Dedekind); the proof for a *general* recursive (and so for an *ordinal* recursive) $f(x)$ was obtained by S. C.

Kleene (see e.g. S. C. Kleene, *Introduction to Metamathematics*, 1953). The formulation of the syntax of arithmetic within arithmetic by numbering symbols, sentences, and proof was also discovered by K. Gödel. (Another method of numbering is described in the Hilbert-Bernays *Grundlagen*.)

The discovery, and proof, that arithmetic is incomplete, and the derivation of the impossibility of a proof of arithmetic's freedom from contradiction without transcending the resources of arithmetic is all the work of K. Gödel ("Über formal unentscheidbare Sätze der *Principia Mathematica* und verwandter Systeme" in *Monatshefte für Math. u. Phys.*, Vol. XXXVIII, 1931; also *On undecidable propositions of formal mathematical systems*, Princeton, 1934). The most detailed and complete account of Gödel's results is also in the Hilbert-Bernays *Grundlagen*.

The impossibility of characterizing the natural numbers by a finite or denumerable infinity of axioms was discovered by Th. Skolem (Norway, 1933; also in *Fundamenta Mathematicæ*, Vol. XXIII, 1934).

Consistency proof

Proofs of the consistency of formalized arithmetic have been given by G. Gentzen (*Math. Annalen*, Vol. CXII, 1936) and W. Ackermann (*ibid.*, Vol. CXVII, 1940); both Gentzen's and Ackermann's proofs appeal to transfinite induction over ordinals up to ε_0 (the smallest ordinal to satisfy $\omega^\omega = \omega$). Gentzen has shown that transfinite induction up to any ordinal $\alpha < \varepsilon_0$ may be proved in system Z, but that transfinite induction up to ε_0 itself is not so provable (as is indirectly established by the conjunction of the Gödel and the Gentzen results).

The decision problem

The first proof of the undecidability of the predicate calculus was obtained by Alonzo Church (*Journal of Symbolic Logic*, Vol. I, 1936). The present proof of the undecidability of arithmetic is due to A. Tarski, A. Mostowski, and R. M. Robinson (*Undecidable Theories*, Amsterdam, 1953). The undecidability of the elementary theory of groups was discovered by Tarski (1946), and the *decidability* of the theory of Abelian Groups is a result obtained by Wanda Szmielew

(X*th International Congress of Philosophy*, Amsterdam, 1948). Tarski has also found a decision method for elementary algebra and geometry (with real number variables, but lacking variables for natural numbers).

The extended predicate calculus

The system described here is due to W. van Orman Quine, and was first published in 1940. This first account was found by J. B. Rosser to admit a contradiction, and the present account (*Mathematical Logic*, revised edition, Harvard, 1951) is protected against this contradiction by restricting bound variables to elements in the membership axioms (a device due to Hao Wang). There are several alternative systems of set theory. The earliest is *Type Theory*, which arranges all objects in a hierarchy of types and is obliged to postulate an axiom of infinity. Another is the Zermelo-Fraenkel system which first introduced the criterion of elementhood for membership of classes; this system was formalized and developed by von Neumann (1926), and reformulated by Bernays (1937). An earlier system of Quine's, known as *New Foundations*, is somewhere between *Type Theory* and the Zermelo-Fraenkel system. The chief difference between the von Neumann-Bernays system and *Mathematical Logic* is in the conditions of elementhood, the former system comprising (approximately) all the classes of the Zermelo system and the latter the more extensive range of classes of *New Foundations*. Wang has recently shown that *Mathematical Logic* is consistent if (the more restrictive) *New Foundations* is consistent.

K. Gödel has proved that the axiom of choice cannot introduce a contradiction into von Neumann-Bernays set-theory if it is otherwise consistent, but E. Specker has recently shown that the axiom of choice is false in *New Foundations*. A brief comparison of the several set theories is given by Hao Wang and Robert McNaughton, *Les Systemes Axiomatiques de la Theorie des Ensembles* (Paris, 1953).

The paradox of the class of subclasses was discovered by Th. Skolem to whom is due also the 'resolution' of the paradox by the relativization of classes. The subject of non-standard models for arithmetic is fully discussed in S. C. Kleene, *Introduction to Metamathematics*.

INDEX

Abelian groups 87, 100
abstraction 57
Ackermann, W. 99, 100
algebra, modern, 43
  elementary 101
alphabet 9
analysis, modern 89
ancestral 95
argument 35, 83
arithmetic 1, 9, 28, 83
  axioms of 44
  elementary 48
  incompleteness of 73, 100
  recursive 60, 82, 89
  undecidability of 85
arithmetical predicate 45, 48
  difference 50
Austin, J. L. 98
axiomatic theory 18
  system 18, 26
axiom of choice 101
axioms 18, 22, 23, 28, 40
  independence of 38
  of arithmetic 44, 74

Bernays, P. 98, 99, 100, 101
Boole, G. 11
bound variable 4, 28
bounded quantifier 52
bracket-free notation 23, 98

calculus, conversion 57
  equation 61
  predicate 28, 44, 67
  sentence 11, 28, 44, 67
canonical form 56
Cantor, G. 78
cardinal number 7
chess 9
choice, axiom of 101
Church, A. 57, 99, 100, 102
class 2, 10, 89
  infinite 89
  logic 8, 89
  null 7, 93
  of all natural numbers 95
  variable 90
commutativity of addition 62
  of multiplication 68
completeness 20, 30, 39, 97, 98
computable functions 83
concatenation 57
conjunction 4, 13, 37, 88
connectives 12, 41, 67
consistent set of statements 39
consistency of sentence logic 21
  of predicate logic 30
constant function 81
contradiction 20, 77, 87
conversion calculus 57
counting 8
course-of-values recursion 53, 76
Curry, H. B. 99

decision procedure 14, 37, 83, 87
  problem 83, 100
Dedekind, R. 99
deduction theorem 19, 31, 42, 69, 88, 98
derivation rule 6, 19, 22, 39
  scheme 61
descriptions operator 94
difference, arithmetical 50
disjunction 13, 37
  inclusive 13
domain 35
  of natural numbers 39, 97
double recursion 50
duality 29

element 90
"Elements" of Euclid 11
elimination rules 25, 33
equalizing rule 61
equation calculus 61
Euclid 11
existence 3, 28
existential quantifier 4, 90
  statement 37, 74
exponential function 50, 60
extended predicate logic 89, 101

factor 47
falsehood 12
Fermat, P. de 83
formal system 61
Fraenkel, A. 101, 102
free variable 4, 28, 61
Frege, G. 2, 98
functions 14, 44
  initial 49
  monotonic 79
  primitive recursive 49
  recursive 48
  representing 14, 17, 51, 85

general recursive function 99
Gentzen, G. 98, 99, 100
geometry, elementary 101

Gödel, K. 38, 99, 100
  numbering 73, 100
Goldbach, C. 83
groups, elementary theory of 87, 100
  Abelian 87

Henkin, L. 99
Herbrand, J. 98
Heyting, A. 98
Hilbert, D. 99, 100

image 95
implication 3
incompleteness of arithmetic 73
  of class logic 97
inequalities 46
independence (of axioms) 21, 98
induction (mathematical) 45, 47, 95
  transfinite 79
inference, natural 25, 33, 98
  rules of 26, 28, 33, 44, 74
initial functions 49
interconvertibility 58
interpretation 30
introduction rules 25, 33
introductory equations 49

Kleene, S. C. 100, 101

λ-conversion 57, 99
λ-definability 59, 83
logic 1
  class 8
  classical 22
  intuitionistic 22
  predicate 8, 61, 83, 88
    extended 89
  sentence 61, 83, 90
  three-valued 17
Łukasiewicz, J. 98

mathematical induction 45, 47, 95
  analysis 99
mathematics 11
McKinsey, J. C. C. 98
McNaughton, R. 101
modern algebra 43
modus ponens 19
monadic predicates 38
money signs 9
monotonic function 79
Mostowski, A. 100
multiple recursion 51
music signs 9
Myhill, J. R. 99

natural inference 25, 33, 98
natural numbers 2, 39, 45, 79, 100
  class of all 95
negation 12
Neumann, J. von 101
normal form 91, 99
n-ply recursive functions 51
numbering, Gödel 73, 100
number 1, 93
  cardinal 7
  Frege-Russell definition of 2, 7, 8, 9, 10, 98
  natural 2, 39, 42
  prime 6
  signs 8
  theory 44
numerals 8, 45, 77

ω-consistency 78
operator, descriptions 94
ordered pair 94
ordinal recursion 51, 83, 99

pair, ordered 94
paradox, class logic 96
parameter 49, 51
Peano, G. 44
Peirce, C. S. 98
Péter, R. 99
Post, E. L. 98
potential 78
powers, relative 93
predecessor function 50
predicate logic (extended) 89
predicate 8
  arithmetical 45
  calculus 28, 44, 61, 67, 74, 99
  logic, completeness of 38, 97, 99
  recursive 51, 69
  satisfiable 35
  valid 35
prime number 6, 47
  relatively 47
primitive recursive function 49, 72, 82, 99
  relation 72, 75
product function 44, 50, 60, 93

quantifier 4
  bounded 52
  existential 90
  universal 90
Quine, W. v. O. 8, 99, 101
quotient 47

recursion, double 50
  course-of-values 53
  multiple 51
  ordinal 51
  primitive 51
  simultaneous 55
  with parameter substitution 54

recursive arithmetic 60, 82, 89, 99
function 48
predicate 51, 69
relation 51, 67, 75
reflexive relation 36
relation, reflexive 36
transitive 36
relative powers 93
remainder 47
representing function 27, 51
restricted variable 34
Robinson, J. 99
Robinson, R. M. 100
Rosser, J. B. 101
rules of derivation 6, 19, 22, 61
elimination 25, 33
inference 26, 28, 33, 44, 74
introduction 25, 33
rules of substitution 61
Russell, Bertrand 2, 98

satisfiable predicate 35
Schröder, E. 98
scope (of quantifier) 28
sentence 11, 18, 29, 61, 74
calculus 11, 28, 44, 67
logic 61, 83, 90, 98
set theory 89, 101
similarity 2
Skolem, T. 99, 100
Smielew, W. 100
Specker, E. 101
stratification 91
substitution axiom 92
formula 66
rules 61
successor function 44, 81, 95
relation 93
sum function 44, 50, 59, 93
syntax 73, 100

tables, truth 12, 16, 17, 22
Tarski, A. 100, 101
tautology 13, 14, 33, 61
term 45
tertium non datur 13, 18, 19, 22
three-valued logic 17, 22
topology 89
transfinite induction 79, 100
ordinal 78
recursion 99
truth 12
tables 12, 16, 17, 22, 98
types 91, 101
type theory 101
typing, potential 91

undecidability of arithmetic 85
of predicate logic 38
unity 7, 47
universality 3, 28
universal quantifier 4, 90
universally valid sentence 13, 18, 26, 31, 36, 88
universal statement 37, 74

valid predicate 35
valid sentence 13
variable 4, 11
bound 28
free 28
restricted 34

Wang, Hao 101
Whitehead, A. N. 3, 98
Wittgenstein, L. 98

Zahlentheorie 48
Zermelo, E. 101
zero 7, 44

and the derivation rule (known as *modus ponens*),

$$p,\ p \to q \vdash q.$$

We require also a rule of substitution that proved sentences are obtained by writing a sentence in place of a variable wherever it occurs in an axiom or proved sentence. To illustrate the operation of the system we give in detail the proof of the *tertium-non-datur*.

1. $\vdash p \to p \vee q$
2. $\vdash p \to p \vee p$
3. $\vdash p \vee p \to p$
4. $\vdash (p \to q) \to (r \vee p \to r \vee q)$
5. $\vdash (p \vee p \to p) \to ((p \to p \vee p) \to (p \to p))$
6. $\vdash (p \to p \vee p) \to (p \to p)$
7. $\vdash p \to p$, i.e. $\neg p \vee p$
8. $\vdash p \vee q \to q \vee p$
9. $\vdash \neg p \vee p \to p \vee \neg p$
10. $\vdash p \vee \neg p$

Commentary on the proof

The first sentence of the proof is axiom (ii); sentence 2 is obtained by substituting p for q in sentence 1; sentence 3 is axiom (i) and sentence 4 is axiom (iv); sentence 5 is obtained by substituting '$p \vee p$' for 'p', 'p' for 'q', and '$\neg p$' for 'r' in 4 (and remembering that $p \to q$ is just an abbreviation for $\neg p \vee q$); sentence 6 follows from sentences 3 and 5, and sentence 7 from sentences 2 and 6; sentence 8 is axiom (iii), sentence 9 is gained from 8 by writing '$\neg p$' for 'p' and 'p' for 'q', and sentence 10 follows from sentences 7 and 9.

By the derivation rule q follows from p and $p \to q$; the converse of this rule, known as the *deduction theorem*, holds in the form: if by adding p to the axioms we can prove q, then $p \to q$ can be proved on axioms (i) to (iv). The deduction theorem is of great importance and is proved by showing that throughout the proof of q we may replace each sentence A by $p \to$ A; the assumption p is replaced by the proved sentence $p \to p$ and the final sentence of the proof is changed from q to $p \to q$.