GÖDEL'S PROOF

by

Ernest Nagel

and

James R. Newman

ROUTLEDGE & KEGAN PAUL LTD

Broadway House, 68-74 Carter Lane

London, E.C.4

I

Introduction

In 1931 there appeared in a German scientific periodical a relatively short paper with the forbidding title "Uber formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme" ("On Formally Undecidable Propositions of Principia Mathematica and Related Systems"). Its author was Kurt Gödel, then a young mathematician of 25 at the University of Vienna and since 1938 a permanent member of the Institute for Advanced Study at Princeton. The paper is a milestone in the history of logic and mathematics. When Harvard University awarded Gödel an honorary degree in 1952, the citation described the work as one of the most important advances in logic in modern times.

At the time of its appearance, however, neither the title of Gödel's paper nor its content was intelligible to most mathematicians. The *Principia Mathematica* mentioned in the title is the monumental three-volume treatise by Alfred North Whitehead and Bertrand Russell on mathematical logic and the foundations of mathematics; and familiarity with that work is not a prerequisite to successful research in most branches of mathematics. Moreover, Gödel's paper deals with a set of questions that has never attracted more than a comparatively small group of students. The reasoning of the proof was so novel at the time of its publication that only those intimately conversant with the technical literature of a highly specialized field could follow the argument with ready comprehension. Nevertheless, the conclusions Gödel established are now widely recognized as being revolutionary in their broad philosophical import. It is the aim of the present essay to make the substance of Gödel's findings and the general character of his proof accessible to the non-specialist.

Gödel's famous paper attacked a central problem in the foundations of mathematics. It will be helpful to give a brief preliminary account of the context in which the problem occurs. Everyone who has been exposed to elementary geometry will doubtless recall that it is taught as a *deductive* discipline. It is not presented as an experimental science whose theorems are to be accepted because they are in agreement with observation. This notion, that a proposition may be established as the conclusion of an explicit *logical proof*, goes back to the ancient Greeks, who discovered what is known as the "axiomatic method" and used it to develop geometry in a systematic fashion. The axiomatic method consists in accepting *without* proof certain propositions as axioms or postulates (e.g., the

axiom that through two points just one straight line can be drawn), and then deriving from the axioms all other propositions of the system as theorems. The axioms constitute the "foundations" of the system; the theorems are the "superstructure," and are obtained from the axioms with the exclusive help of principles of logic.

The axiomatic development of geometry made a powerful impression upon thinkers throughout the ages; for the relatively small number of axioms carry the whole weight of the inexhaustibly numerous propositions derivable from them. Moreover, if in some way the truth of the axioms can be established—and, indeed, for some two thousand years most students believed without question that they are true of space—both the truth and the mutual consistency of all the theorems are automatically guaranteed. For these reasons the axiomatic form of geometry appeared to many generations of outstanding thinkers as the model of scientific knowledge at its best. It was natural to ask, therefore, whether other branches of thought besides geometry can be placed upon a secure axiomatic foundation. However, although certain parts of physics were given an axiomatic formulation in antiquity (e.g., by Archimedes), until modern times geometry was the only branch of mathematics that had what most students considered a sound axiomatic basis.

But within the past two centuries the axiomatic method has come to be exploited with increasing power and vigor. New as well as old branches of mathematics, including the familiar arithmetic of cardinal (or "whole") numbers, were supplied with what appeared to be adequate sets of axioms. A climate of opinion was thus generated in which it was tacitly assumed that each sector of mathematical thought can be supplied with a set of axioms sufficient for developing systematically the endless totality of true propositions about the given area of inquiry.

Gödel's paper showed that this assumption is untenable. He presented mathematicians with the astounding and melancholy conclusion that the axiomatic method has certain inherent limitations, which rule out the possibility that even the ordinary arithmetic of the integers can ever be fully axiomatized. What is more, he proved that it is impossible to establish the internal logical consistency of a very large class of deductive systems—elementary arithmetic, for example—unless one adopts principles of reasoning so complex that their internal consistency is as open to doubt as that of the systems themselves. In the light of these conclusions, no final systematization of many important areas of mathematics is attainable, and no absolutely impeccable guarantee can be given that many significant branches of mathematical thought are entirely free from internal contradiction.

Gödel's findings thus undermined deeply rooted preconceptions and demolished ancient hopes that were being freshly nourished by research on the foundations of mathematics. But his paper was not altogether negative. It introduced into the study of foundation questions a new technique of analysis comparable in its nature and fertility with the algebraic method that René Descartes introduced into geometry. This technique suggested and initiated new problems for logical and mathematical investigation. It provoked a reappraisal, still under way, of widely held philosophies of mathematics, and of philosophies of knowledge in general.

The details of Gödel's proofs in his epoch-making paper are too difficult to follow without considerable mathematical training. But the basic structure of his demonstrations and the core of his conclusions can be made intelligible to readers with very limited mathematical and logical preparation. To achieve such an understanding, the reader may find useful a brief account of certain relevant developments in the history of mathematics and of modern formal logic. The next four sections of this essay are devoted to this survey.

II

The Problem of Consistency

The nineteenth century witnessed a tremendous expansion and intensification of mathematical research. Many fundamental problems that had long withstood the best efforts of earlier thinkers were solved; new departments of mathematical study were created; and in various branches of the discipline new foundations were laid, or old ones entirely recast with the help of more precise techniques of analysis. To illustrate: the Greeks had proposed three problems in elementary geometry: with compass and straight-edge to trisect any angle, to construct a cube with a volume twice the volume of a given cube, and to construct a square equal in area to that of a given circle. For more than 2,000 years unsuccessful attempts were made to solve these problems; at last, in the nineteenth century it was proved that the desired constructions are logically impossible. There was, moreover, a valuable by-product of these labors. Since the solutions depend essentially upon determining the kind of roots that satisfy certain equations, concern with the celebrated

exercises set in antiquity stimulated profound investigations into the nature of number and the structure of the number continuum. Rigorous definitions were eventually supplied for negative, complex, and irrational numbers; a logical basis was constructed for the real number system; and a new branch of mathematics, the theory of infinite numbers, was founded.

But perhaps the most significant development in its long-range effects upon subsequent mathematical history was the solution of another problem that the Greeks raised without answering. One of the axioms Euclid used in systematizing geometry has to do with parallels. The axiom he adopted is logically equivalent to (though not identical with) the assumption that through a point outside a given line only one parallel to the line can be drawn. For various reasons, this axiom did not appear "self-evident" to the ancients. They sought, therefore, to deduce it from the other Euclidean axioms, which they regarded as clearly self-evident.[1] Can such a proof of the parallel axiom be

---

[1] The chief reason for this alleged lack of self-evidence seems to have been the fact that the parallel axiom makes an assertion about *infinitely remote* regions of space. Euclid defines parallel lines as straight lines in a plane that, "being produced indefinitely in both directions," do not meet. Accordingly, to say that two lines are parallel is to make the claim that the two lines will not meet even "at infinity." But the ancients were familiar with lines that, though they do not intersect each other in any finite region of the plane, do meet "at infinity." Such lines are said to be "asymptotic." Thus, a hyperbola is asymptotic to its axes. It was therefore not in-

given? Generations of mathematicians struggled with this question, without avail. But repeated failure to construct a proof does not mean that none can be found any more than repeated failure to find a cure for the common cold establishes beyond doubt that mankind will forever suffer from running noses. It was not until the nineteenth century, chiefly through the work of Gauss, Bolyai, Lobachevsky, and Riemann, that the *impossibility* of deducing the parallel axiom from the others was demonstrated. This outcome was of the greatest intellectual importance. In the first place, it called attention in a most impressive way to the fact that a *proof* can be given of the *impossibility of proving* certain propositions within a given system. As we shall see, Gödel's paper is a proof of the impossibility of demonstrating certain important propositions in arithmetic. In the second place, the resolution of the parallel axiom question forced the realization that Euclid is not the last word on the subject of geometry, since new systems of geometry can be constructed by using a number of axioms different from, and incompatible with, those adopted by Euclid. In particular, as is well known, immensely interesting and fruitful results are obtained when Euclid's parallel axiom is replaced by the assumption that more than one parallel can be drawn to a given line through a given point, or, alternatively, by the assumption that no parallels can

tuitively evident to the ancient geometers that from a point outside a given straight line only one straight line can be drawn that will not meet the given line even at infinity.

be drawn. The traditional belief that the axioms of geometry (or, for that matter, the axioms of any discipline) can be established by their apparent self-evidence was thus radically undermined. Moreover, it gradually became clear that the proper business of the pure mathematician is to *derive theorems from postulated assumptions,* and that it is not his concern as a mathematician to decide whether the axioms he assumes are actually true. And, finally, these successful modifications of orthodox geometry stimulated the revision and completion of the axiomatic bases for many other mathematical systems. Axiomatic foundations were eventually supplied for fields of inquiry that had hitherto been cultivated only in a more or less intuitive manner. (See Appendix, no. 1.)

The over-all conclusion that emerged from these critical studies of the foundations of mathematics is that the age-old conception of mathematics as "the science of quantity" is both inadequate and misleading. For it became evident that mathematics is simply the discipline *par excellence* that draws the conclusions logically implied by any given set of axioms or postulates. In fact, it came to be acknowledged that the validity of a mathematical inference in no sense depends upon any special meaning that may be associated with the terms or expressions contained in the postulates. Mathematics was thus recognized to be much more abstract and formal than had been traditionally supposed: more abstract, because mathematical statements can be construed in principle to be about anything whatsoever rather than about some inherently circumscribed set of objects or traits of objects; and more formal, because the validity of mathematical demonstrations is grounded in the structure of statements, rather than in the nature of a particular subject matter. The postulates of any branch of demonstrative mathematics are not inherently about space, quantity, apples, angles, or budgets; and any special meaning that may be associated with the terms (or "descriptive predicates") in the postulates plays no essential role in the process of deriving theorems. We repeat that the sole question confronting the pure mathematician (as distinct from the scientist who employs mathematics in investigating a special subject matter) is not whether the postulates he assumes or the conclusions he deduces from them are true, but whether the alleged conclusions are in fact the *necessary logical consequences* of the initial assumptions.

Take this example. Among the undefined (or "primitive") terms employed by the influential German mathematician David Hilbert in his famous axiomatization of geometry (first published in 1899) are 'point', 'line', 'lies on', and 'between'. We may grant that the customary meanings connected with these expressions play a role in the process of discovering and learning theorems. Since the meanings are familiar, we feel we understand their various interrelations, and they motivate the formulation and selection of axioms; moreover, they suggest and facilitate the formulation of the statements we hope to establish

as theorems. Yet, as Hilbert plainly states, insofar as we are concerned with the primary mathematical task of exploring the purely logical relations of dependence between statements, the familiar connotations of the primitive terms are to be ignored, and the sole "meanings" that are to be associated with them are those assigned by the axioms into which they enter.[2] This is the point of Russell's famous epigram: pure mathematics is the subject in which we do not know what we are talking about, or whether what we are saying is true.

A land of rigorous abstraction, empty of all familiar landmarks, is certainly not easy to get around in. But it offers compensations in the form of a new freedom of movement and fresh vistas. The intensified formalization of mathematics emancipated men's minds from the restrictions that the customary interpretation of expressions placed on the construction of novel systems of postulates. New kinds of algebras and geometries were developed which marked significant departures from the mathematics of tradition. As the meanings of certain terms became more general, their use became broader and the inferences that could be drawn from them less confined. Formalization led to a great variety of systems of considerable mathematical interest and

---

[2] In more technical language, the primitive terms are "implicitly" defined by the axioms, and whatever is not covered by the implicit definitions is irrelevant to the demonstration of theorems.

value. Some of these systems, it must be admitted, did not lend themselves to interpretations as obviously intuitive (i.e., commonsensical) as those of Euclidean geometry or arithmetic, but this fact caused no alarm. Intuition, for one thing, is an elastic faculty: our children will probably have no difficulty in accepting as intuitively obvious the paradoxes of relativity, just as we do not boggle at ideas that were regarded as wholly unintuitive a couple of generations ago. Moreover, as we all know, intuition is not a safe guide: it cannot properly be used as a criterion of either truth or fruitfulness in scientific explorations.

However, the increased abstractness of mathematics raised a more serious problem. It turned on the question whether a given set of postulates serving as foundation of a system is internally consistent, so that no mutually contradictory theorems can be deduced from the postulates. The problem does not seem pressing when a set of axioms is taken to be about a definite and familiar domain of objects; for then it is not only significant to ask, but it may be possible to ascertain, whether the axioms are indeed true of these objects. Since the Euclidean axioms were generally supposed to be true statements about space (or objects in space), no mathematician prior to the nineteenth century ever considered the question whether a pair of contradictory theorems might some day be deduced from the axioms. The basis for this confidence in the consistency of Euclidean geometry is the sound principle that logically incompatible statements cannot be simultaneously true; accordingly, if a set of statements is true (and this was assumed of the Euclidean axioms), these statements are mutually consistent.

The non-Euclidean geometries were clearly in a different category. Their axioms were initially regarded as being plainly false of space, and, for that matter, doubtfully true of anything; thus the problem of establishing the internal consistency of non-Euclidean systems was recognized to be both formidable and critical. In Riemannian geometry, for example, Euclid's parallel postulate is replaced by the assumption that through a given point outside a line *no* parallel to it can be drawn. Now suppose the question: Is the Riemannian set of postulates consistent? The postulates are apparently not true of the space of ordinary experience. How, then, is their consistency to be shown? How can one prove they will not lead to contradictory theorems? Obviously the question is not settled by the fact that the theorems already deduced do not contradict each other—for the possibility remains that the very next theorem to be deduced may upset the apple cart. But, until the question is settled, one cannot be certain that Riemannian geometry is a true alternative to the Euclidean system, i.e., equally valid mathematically. The very possibility of non-Euclidean geometries was thus contingent on the resolution of this problem.

A general method for solving it was devised. The underlying idea is to find a "model" (or "interpretation") for the abstract postulates of a system, so that each postulate is converted into a true statement about the model. In the case of Euclidean geometry, as we have noted, the model was ordinary space. The method was used to find other models, the elements of which could serve as crutches for determining the consistency of abstract postulates. The procedure goes something like this. Let us understand by the word 'class' a collection or aggregate of distinguishable elements, each of which is called a member of the class. Thus, the class of prime numbers less than 10 is the collection whose members are 2, 3, 5, and 7. Suppose the following set of postulates concerning two classes K and L, whose special nature is left undetermined except as "implicitly" defined by the postulates:

1. Any two members of K are contained in just one member of L.
2. No member of K is contained in more than two members of L.
3. The members of K are not all contained in a single member of L.
4. Any two members of L contain just one member of K.
5. No member of L contains more than two members of K.

From this small set we can derive, by using customary rules of inference, a number of theorems. For example, it can be shown that K contains just three members. But is the set consistent, so that mutually contradictory theorems can never be derived from

it? The question can be answered readily with the help of the following model:

> Let K be the class of points consisting of the vertices of a triangle, and L the class of lines made up of its sides; and let us understand the phrase 'a member of K is contained in a member of L' to mean that a point which is a vertex lies on a line which is a side. Each of the five abstract postulates is then converted into a true statement. For instance, the first postulate asserts that any two points which are vertices of the triangle lie on just one line which is a side. (See Fig. 1.) In this way the set of postulates is proved to be consistent.

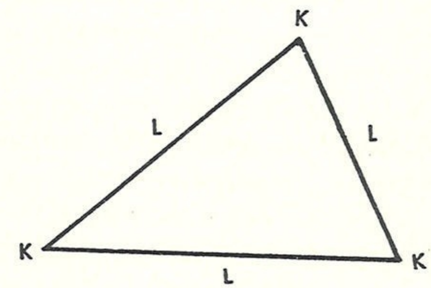

Fig. 1.

Model for a set of postulates about two classes, K and L, is a triangle whose vertices are the members of K and whose sides are the members of L. The geometrical model shows that the postulates are consistent.

The consistency of plane Riemannian geometry can also, ostensibly, be established by a model embodying the postulates. We may interpret the expression 'plane' in the Riemannian axioms to signify the surface of a Euclidean sphere, the expression 'point' a point on this surface, the expression 'straight line' an arc of a great circle on this surface, and so on. Each Riemannian postulate is then converted into a theorem of Euclid. For example, on this interpretation the Riemannian parallel postulate reads: Through a point on the surface of a sphere, no arc of a great circle can be drawn parallel to a given arc of a great circle. (See Fig. 2.)

At first glance this proof of the consistency of Riemannian geometry may seem conclusive. But a closer look is disconcerting. For a sharp eye will discern that the problem has not been solved; it has merely been shifted to another domain. The proof attempts to settle the consistency of Riemannian geometry by appealing to the consistency of Euclidean geometry. What emerges, then, is only this: Riemannian geometry is consistent if Euclidean geometry is consistent. The authority of Euclid is thus invoked to demonstrate the consistency of a system which challenges the exclusive validity of Euclid. The inescapable question is: Are the axioms of the Euclidean system itself consistent?

An answer to this question, hallowed, as we have noted, by a long tradition, is that the Euclidean axioms are true and are therefore consistent. This answer is

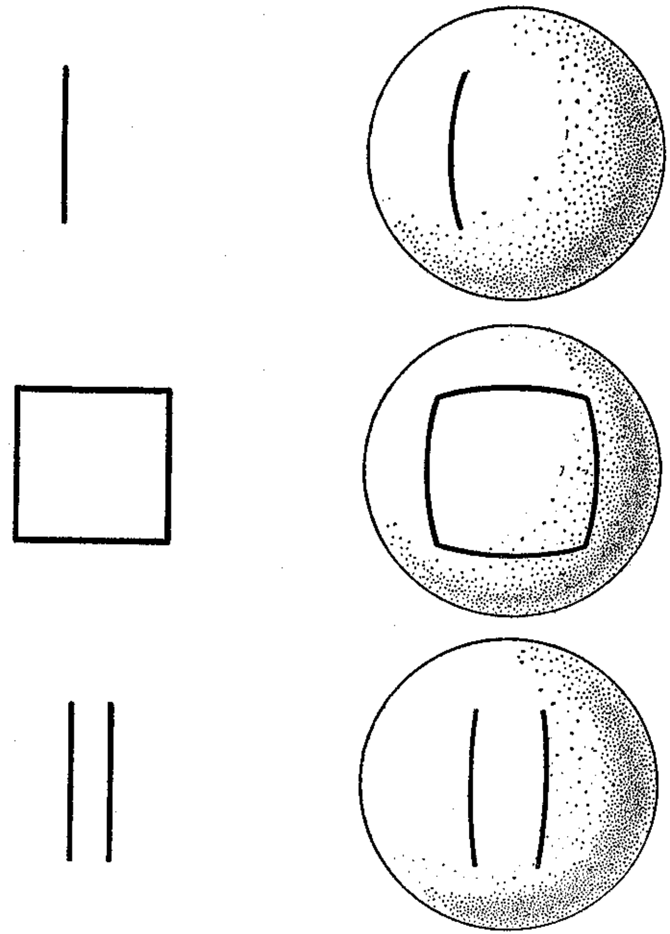

Fig. 2

The non-Euclidean geometry of Bernhard Riemann can be represented by a Euclidean model. The Riemannian plane becomes the surface of a Euclidean sphere, points on the plane become points on this surface, straight lines in the plane become great circles. Thus, a portion of the Riemannian plane bounded by segments of straight lines is depicted as a portion of the sphere bounded by parts of great circles (*center*). Two line segments in the Riemannian plane are two segments of great circles on the Euclidean sphere (*bottom*), and these, if extended, indeed intersect, thus contradicting the parallel postulate.

no longer regarded as acceptable; we shall return to it presently and explain why it is unsatisfactory. Another answer is that the axioms jibe with our actual, though limited, experience of space and that we are justified in extrapolating from the small to the universal. But, although much inductive evidence can be adduced to support this claim, our best proof would be logically incomplete. For even if all the observed facts are in agreement with the axioms, the possibility is open that a hitherto unobserved fact may contradict them and so destroy their title to universality. Inductive considerations can show no more than that the axioms are plausible or probably true.

Hilbert tried yet another route to the top. The clue to his way lay in Cartesian coordinate geometry. In his interpretation Euclid's axioms were simply transformed into algebraic truths. For instance, in the axioms for plane geometry, construe the expression 'point' to signify a pair of numbers, the expression 'straight line' the (linear) relation between numbers expressed by a first degree equation with two unknowns, the expression 'circle' the relation between numbers expressed by a quadratic equation of a certain form, and so on. The geometric statement that two distinct points uniquely determine a straight line is then transformed into the algebraic truth that two distinct pairs of numbers uniquely determine a linear relation; the geometric theorem that a straight line intersects a circle in at most two points, into the algebraic theorem that a pair of simultaneous equations in two unknowns (one of which is linear and the

other quadratic of a certain type) determine at most two pairs of real numbers; and so on. In brief, the consistency of the Euclidean postulates is established by showing that they are satisfied by an algebraic model. This method of establishing consistency is powerful and effective. Yet it, too, is vulnerable to the objection already set forth. For, again, a problem in one domain is resolved by transferring it to another. Hilbert's argument for the consistency of his geometric postulates shows that if algebra is consistent, so is his geometric system. The proof is clearly relative to the assumed consistency of another system and is not an "absolute" proof.

In the various attempts to solve the problem of consistency there is one persistent source of difficulty. It lies in the fact that the axioms are interpreted by models composed of an infinite number of elements. This makes it impossible to encompass the models in a finite number of observations; hence the truth of the axioms themselves is subject to doubt. In the inductive argument for the truth of Euclidean geometry, a finite number of observed facts about space are presumably in agreement with the axioms. But the conclusion that the argument seeks to establish involves an extrapolation from a finite to an infinite set of data. How can we justify this jump? On the other hand, the difficulty is minimized, if not completely eliminated, where an appropriate model can be devised that contains only a finite number of elements. The triangle model used to show the consistency of the five abstract postulates for the classes K and L is finite; and it is comparatively simple to determine by actual inspection whether all the elements in the model actually satisfy the postulates, and thus whether they are true (and hence consistent). To illustrate: by examining in turn all the vertices of the model triangle, one can learn whether any two of them lie on just one side—so that the first postulate is established as true. Since all the elements of the model, as well as the relevant relations among them, are open to direct and exhaustive inspection, and since the likelihood of mistakes occurring in inspecting them is practically nil, the consistency of the postulates in this case is not a matter for genuine doubt.

Unfortunately, most of the postulate systems that constitute the foundations of important branches of mathematics cannot be mirrored in finite models. Consider the postulate in elementary arithmetic which asserts that every integer has an immediate successor differing from any preceding integer. It is evident that the model needed to test the set to which this postulate belongs cannot be finite, but must contain an infinity of elements. It follows that the truth (and so the consistency) of the set cannot be established by an exhaustive inspection of a limited number of elements. Apparently we have reached an impasse. Finite models suffice, in principle, to establish the consistency of certain sets of postulates; but these are of slight mathematical importance. Non-finite models, necessary for the interpretation of most postulate systems of mathematical significance, can be described only in general terms; and we cannot conclude as a matter of course

that the descriptions are free from concealed contradictions.

It is tempting to suggest at this point that we can be sure of the consistency of formulations in which non-finite models are described if the basic notions employed are transparently "clear" and "distinct." But the history of thought has not dealt kindly with the doctrine of clear and distinct ideas, or with the doctrine of intuitive knowledge implicit in the suggestion. In certain areas of mathematical research in which assumptions about infinite collections play central roles, radical contradictions have turned up, in spite of the intuitive clarity of the notions involved in the assumptions and despite the seemingly consistent character of the intellectual constructions performed. Such contradictions (technically referred to as "antinomies") have emerged in the theory of infinite numbers, developed by Georg Cantor in the nineteenth century; and the occurrence of these contradictions has made plain that the apparent clarity of even such an elementary notion as that of *class* (or *aggregate*) does not guarantee the consistency of any particular system built on it. Since the mathematical theory of classes, which deals with the properties and relations of aggregates or collections of elements, is often adopted as the foundation for other branches of mathematics, and in particular for elementary arithmetic, it is pertinent to ask whether contradictions similar to those encountered in the theory of infinite classes infect the formulations of other parts of mathematics.

In point of fact, Bertrand Russell constructed a contradiction within the framework of elementary logic itself that is precisely analogous to the contradiction first developed in the Cantorian theory of infinite classes. Russell's antinomy can be stated as follows. Classes seem to be of two kinds: those which do not contain themselves as members, and those which do. A class will be called "normal" if, and only if, it does not contain itself as a member; otherwise it will be called "non-normal." An example of a normal class is the class of mathematicians, for patently the class itself is not a mathematician and is therefore not a member of itself. An example of a non-normal class is the class of all thinkable things; for the class of all thinkable things is itself thinkable and is therefore a member of itself. Let 'N' by definition stand for the class of *all* normal classes. We ask whether N itself is a normal class. If N is normal, it is a member of itself (for by definition N contains all normal classes); but, in that case, N is non-normal, because by definition a class that contains itself as a member is non-normal. On the other hand, if N is non-normal, it is a member of itself (by definition of non-normal); but, in that case, N is normal, because by definition the members of N are normal classes. In short, N is normal if, and only if, N is non-normal. It follows that the statement 'N is normal' is both true and false. This fatal contradiction results from an uncritical use of the apparently pellucid notion of class. Other paradoxes were found later, each of them constructed by means of familiar and seemingly cogent modes of reasoning. Mathematicians came to realize that in developing consistent systems familiarity and intuitive clarity are weak reeds to lean on.

We have seen the importance of the problem of consistency, and we have acquainted ourselves with the classically standard method for solving it with the help of models. It has been shown that in most instances the problem requires the use of a non-finite model, the description of which may itself conceal inconsistencies. We must conclude that, while the model method is an invaluable mathematical tool, it does not supply a final answer to the problem it was designed to solve.

VII

Gödel's Proofs

Gödel's paper is difficult. Forty-six preliminary definitions, together with several important preliminary theorems, must be mastered before the main results are reached. We shall take a much easier road; nevertheless, it should afford the reader glimpses of the ascent and of the crowning structure.

A *Gödel numbering*

Gödel described a formalized calculus within which all the customary arithmetical notations can be expressed and familiar arithmetical relations established.[15] The formulas of the calculus are constructed out of a class of elementary signs, which constitute the fundamental vocabulary. A set of primitive formulas (or axioms) are the underpinning, and the theorems of the calculus are formulas derivable from the axioms with the help

[15] He used an adaptation of the system developed in *Principia Mathematica*. But any calculus within which the cardinal number system can be constructed would have served his purpose.

of a carefully enumerated set of Transformation Rules (or rules of inference).

Gödel first showed that it is possible to assign a *unique number* to each elementary sign, each formula (or sequence of signs), and each proof (or finite sequence of formulas). This number, which serves as a distinctive tag or label, is called the "Gödel number" of the sign, formula, or proof.[16]

The elementary signs belonging to the fundamental vocabulary are of two kinds: the constant signs and the variables. We shall assume that there are exactly ten constant signs,[17] to which the integers from 1 to 10 are attached as Gödel numbers. Most of these signs are already known to the reader: '$\sim$' (short for 'not'); 'V' (short for 'or'); '$\supset$' (short for 'if . . . then . . .'); '=' (short for 'equals'); '0' (the numeral for the number zero); and three signs of punctuation, namely, the left parenthesis '(', the right parenthesis ')', and the comma ','. In addition, two other signs will be used: the inverted letter '$\exists$', which may be read as 'there is' and which occurs in "existential quantifiers"; and the

[16] There are many alternative ways of assigning Gödel numbers, and it is immaterial to the main argument which is adopted. We give a concrete example of how the numbers can be assigned to help the reader follow the discussion. The method of numbering used in the text was employed by Gödel in his 1931 paper.

[17] The number of constant signs depends on how the formal calculus is set up. Gödel in his paper used only seven constant signs. The text uses ten in order to avoid certain complexities in the exposition.

lower-case 's', which is attached to numerical expressions to designate the immediate successor of a number. To illustrate: the formula '$(\exists x)(x = s0)$' may be read 'There is an x such that x is the immediate successor of 0'. The table below displays the ten constant signs, states the Gödel number associated with each, and indicates the usual meanings of the signs.

Constant Signs	Gödel Number	Meaning
$\sim$	1	not
$\vee$	2	or
$\supset$	3	If . . . then
$\exists$	4	There is an . . .
$=$	5	equals
0	6	zero
s	7	The immediate successor of
(	8	punctuation mark
)	9	punctuation mark
,	10	punctuation mark

TABLE 2

Besides the constant elementary signs, three kinds of variables appear in the fundamental vocabulary of the calculus: the *numerical variables* 'x', 'y', 'z', etc., for which numerals and numerical expressions may be substituted; the *sentential variables* 'p', 'q', 'r', etc., for which formulas (sentences) may be substituted; and the *predicate variables* 'P', 'Q', 'R', etc., for which predicates, such as 'Prime' or 'Greater than', may be substituted. The variables are assigned Gödel numbers in accordance with the following rules: associate (i) with each distinct numerical variable a distinct prime number greater than 10; (ii) with each distinct sentential variable the square of a prime number greater than 10; and (iii) with each distinct predicate variable the cube of a prime greater than 10. The accompanying table illustrates the use of these rules to specify the Gödel numbers of a few variables.

Numerical Variable	Gödel Number	A Possible Substitution Instance
x	11	0
y	13	s0
z	17	y

Numerical variables are associated with prime numbers greater than 10.

Sentential Variable	Gödel Number	A Possible Substitution Instance
p	11^2	$0 = 0$
q	13^2	$(\exists x)(x = sy)$
r	17^2	$p \supset q$

Sentential variables are associated with the squares of prime numbers greater than 10.

Predicate Variable	Gödel Number	A Possible Substitution Instance
P	11^3	Prime
Q	13^3	Composite
R	17^3	Greater than

Predicate Variables are associated with the cubes of prime numbers greater than 10.

TABLE 3

Consider next a formula of the system, for example, '$(\exists x)(x = sy)$'. (Literally translated, this reads: 'There is an *x* such that *x* is the immediate successor of *y*', and says, in effect, that every number has an immediate successor.) The numbers associated with its ten constituent elementary signs are, respectively, 8, 4, 11, 9, 8, 11, 5, 7, 13, 9. We show this schematically below:

(	∃	*x*	)	(	*x*	=	s	*y*	)
↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
8	4	11	9	8	11	5	7	13	9

It is desirable, however, to assign a single number to the formula rather than a set of numbers. This can be done easily. We agree to associate with the formula the unique number that is the product of the first ten primes in order of magnitude, each prime being raised to a power equal to the Gödel number of the corresponding elementary sign. The above formula is accordingly associated with the number

$$2^8 \times 3^4 \times 5^{11} \times 7^9 \times 11^8 \times 13^{11} \times 17^5 \times 19^7 \times 23^{13} \times 29^9;$$

let us refer to this number as *m*. In a similar fashion, a unique number, the product of as many primes as there are signs (each prime being raised to a power equal to the Gödel number of the corresponding sign), can be assigned to every finite sequence of elementary signs and, in particular, to every formula.[18]

Consider, finally, a sequence of formulas, such as may occur in some proof, e.g., the sequence:

$$(\exists x)(x = sy)$$
$$(\exists x)(x = s0)$$

The second formula when translated reads '0 has an

[18] Signs may occur in the calculus which do not appear in the fundamental vocabulary; these are introduced by defining them with the help of the vocabulary signs. For example, the sign '·', the sentential connective used as an abbreviation for 'and', can be defined in context as follows: '$p \cdot q$' is short for '$\sim(\sim p \vee \sim q)$'. What Gödel number is assigned to a defined sign? The answer is obvious if we notice that expressions containing defined signs can be eliminated in favor of their defining equivalents; and it is clear that a Gödel number can be determined for the transformed expressions. Accordingly, the Gödel number of the formula '$p \cdot q$' is the Gödel number of the formula '$\sim(\sim p \vee \sim q)$'. Similarly, the various numerals can be introduced by definition as follows: '1' as short for 's0', '2' as short for 'ss0', '3' as short for 'sss0', and so on. To obtain the Gödel number for the formula '$\sim(2 = 3)$', we eliminate the defined signs, thus obtaining the formula '$\sim(ss0 = sss0)$', and determine its Gödel number pursuant to the rules stated in the text.

immediate successor'; it is derivable from the first by substituting the numeral '0' for the numerical variable 'y'.[19] We have already determined the Gödel number of the first formula: it is m; and suppose that n is the Gödel number of the second formula. As before, it is convenient to have a single number as a tag for the sequence. We agree therefore to associate with it the number which is the product of the first two primes in order of magnitude (i.e., the primes 2 and 3), each prime being raised to a power equal to the Gödel number of the corresponding formula in the sequence. If we call this number k, we can write $k = 2^m \times 3^n$. By applying this compact procedure we can obtain a number for each sequence of formulas. In sum, every expression in the system, whether an elementary sign, a sequence of signs, or a sequence of sequences, can be assigned a unique Gödel number.

What has been done so far is to establish a method for completely "arithmetizing" the formal calculus. The method is essentially a set of directions for setting up a one-to-one correspondence between the expressions in the calculus and a certain subset of the integers.[20] Once an expression is given, the Gödel number uniquely corresponding to it can be calculated. But this is only half the story. Once a number is given, we can determine whether it is a Gödel number, and, if it is, the expression it represents can be exactly analyzed or "retrieved." If a given number is less than or equal to 10, it is the Gödel number of an elementary constant sign. The sign can be identified. If the number is greater than 10, it can be decomposed into its prime factors in just one way (as we know from a famous theorem of arithmetic).[21] If it is a prime greater than 10, or the second or third power of such a prime, it is the Gödel number of an identifiable variable. If it is the product of successive primes, each raised to some power, it may be the Gödel number either of a formula

[19] The reader will recall that we defined a proof as a finite sequence of formulas, each of which either is an axiom or can be derived from preceding formulas in the sequence with the help of the Transformation Rules. By this definition the above sequence is not a proof, since the first formula is not an axiom and its derivation from the axioms is not shown: the sequence is only a segment of a proof. It would take too long to write out a full example of a proof, and for illustrative purposes the above sequence will suffice.

[20] Not every integer is a Gödel number. Consider, for example, the number 100. 100 is greater than 10, and therefore cannot be the Gödel number of an elementary constant sign; and since it is neither a prime greater than 10, nor the square nor the cube of such a prime, it cannot be the Gödel number of a variable. On decomposing 100 into its prime factors, we find that it is equal to $2^2 \times 5^2$; and the prime number 3 does not appear as a factor in the decomposition, but is skipped. According to the rules laid down, however, the Gödel number of a formula (or of a sequence of formulas) must be the product of *successive* primes, each raised to some power. The number 100 does not satisfy this condition. In short, 100 cannot be assigned to constant signs, variables, or formulas; hence it is not a Gödel number.

[21] This theorem is known as the fundamental theorem of arithmetic. It says that if an integer is composite (i.e., not a prime) it has a unique decomposition into prime factors.

or of a sequence of formulas. In this case the expression to which it corresponds can be exactly determined. Following this program, we can take any given number apart, as if it were a machine, discover how it is constructed and what goes into it; and since each of its elements corresponds to an element of the expression it represents, we can reconstitute the expression, analyze its structure, and the like. Table 4 illustrates for a given number how we can ascertain whether it is a Gödel number and, if so, what expression it symbolizes.

A	243,000,000
B	$64 \times 243 \times 15{,}625$
C	$2^6 \times 3^5 \times 5^6$
D	6 5 6 ↓ ↓ ↓ 0 = 0
E	0 = 0

The arithmetical formula 'zero equals zero' has the Gödel number 243 million. Reading down from A to E, the illustration shows how the number is translated into the expression it represents; reading up, how the number for the formula is derived.

TABLE 4

B *The arithmetization of meta-mathematics*

Gödel's next step is an ingenious application of mapping. He showed that all meta-mathematical state-

ments about the structural properties of expressions in the calculus can be adequately *mirrored* within the calculus itself. The basic idea underlying his procedure is this: Since every expression in the calculus is associated with a (Gödel) number, a meta-mathematical statement about expressions and their relations to one another may be construed as a statement about the corresponding (Gödel) numbers and their arithmetical relations to one another. In this way meta-mathematics becomes completely "arithmetized." To take a trivial analogue: customers in a busy supermarket are often given, when they enter, tickets on which are printed numbers whose order determines the order in which the customers are to be waited on at the meat counter. By inspecting the numbers it is easy to tell how many persons have been served, how many are waiting, who precedes whom, and by how many customers, and so on. If, for example, Mrs. Smith has number 37, and Mrs. Brown number 53, instead of explaining to Mrs. Brown that she has to wait her turn after Mrs. Smith, it suffices to point out that 37 is less than 53.

As in the supermarket, so in meta-mathematics. Each meta-mathematical statement is represented by a unique formula within arithmetic; and the relations of logical dependence between meta-mathematical statements are fully reflected in the numerical relations of dependence between their corresponding arithmetical formulas. Once again mapping facilitates an inquiry into structure. The exploration of meta-mathematical questions can be pursued by investigating the arithmetical properties and relations of certain integers.

We illustrate these general remarks by an elementary example. Consider the first axiom of the sentential calculus, which also happens to be an axiom in the formal system under discussion: '$(p \vee p) \supset p$'. Its Gödel number is $2^8 \times 3^{11^2} \times 5^2 \times 7^{11^2} \times 11^9 \times 13^3 \times 17^{11^2}$, which we shall designate by the letter '*a*'. Consider also the formula: '$(p \vee p)$', whose Gödel number is $2^8 \times 3^{11^2} \times 5^2 \times 7^{11^2} \times 11^9$; we shall designate it by the letter '*b*'. We now assert the meta-mathematical statement that the formula '$(p \vee p)$' is an initial part of the axiom. To what arithmetical formula in the formal system does this statement correspond? It is evident that the smaller formula '$(p \vee p)$' can be an initial part of the larger formula which is the axiom if, and only if, the (Gödel) number *b*, representing the former, is a factor of the (Gödel) number *a*, representing the latter. On the assumption that the expression 'factor of' is suitably defined in the formalized arithmetical system, the arithmetical formula which uniquely corresponds to the above meta-mathematical statement is: '*b* is a factor of *a*'. Moreover, if this formula is true, i.e., if *b* is a factor of *a*, then it is true that '$(p \vee p)$' is an initial part of '$(p \vee p) \supset p$'.

Let us fix attention on the meta-mathematical statement: 'The sequence of formulas with Gödel number *x* is a proof of the formula with Gödel number *z*'. This statement is represented (mirrored) by a definite formula in the arithmetical calculus which expresses a

purely arithmetical relation between *x* and *z*. (We can gain some notion of the complexity of this relation by recalling the example used above, in which the Gödel number $k = 2^m \times 3^n$ was assigned to the (fragment of a) proof whose conclusion has the Gödel number *n*. A little reflection shows that there is here a definite, though by no means simple, arithmetical relation between *k*, the Gödel number of the proof, and *n*, the Gödel number of the conclusion.) We write this relation between *x* and *z* as the formula 'Dem (*x*, *z*)', to remind ourselves of the meta-mathematical statement to which it corresponds (i.e., of the meta-mathematical statement 'The sequence of formulas with Gödel number *x* is a proof (or demonstration) of the formula with Gödel number *z*').[22] We now ask the reader to observe that a meta-mathematical statement which says that a certain sequence of formulas is a proof for a given formula is *true*, if, and only if, the Gödel number of the alleged proof stands to the Gödel number of the conclusion in the arithmetical relation here designated by 'Dem'. Accordingly, to establish the truth or falsity of the meta-mathematical statement under discussion, we need concern ourselves only with the ques-

[22] The reader must keep clearly in mind that, though 'Dem (*x*, *z*)' represents the meta-mathematical statement, the formula itself belongs to the arithmetical calculus. The formula could be written in more customary notation as '$f(x, z) = 0$', where the letter '*f*' denotes a complex set of arithmetical operations upon numbers. But this more customary notation does not immediately suggest the meta-mathematical interpretation of the formula.

tion whether the relation Dem holds between two numbers. Conversely, we can establish that the arithmetical relation holds between a pair of numbers by showing that the meta-mathematical statement mirrored by this relation between the numbers is true. Similarly, the meta-mathematical statement 'The sequence of formulas with the Gödel number x is *not* a proof for the formula with the Gödel number z' is represented by a definite formula in the formalized arithmetical system. This formula is the formal contradictory of 'Dem (x, z)', namely, '$\sim$ Dem (x, z)'.

One additional bit of special notation is needed for stating the crux of Gödel's argument. Begin with an example. The formula '$(\exists x)(x = sy)$' has m for its Gödel number (see page 73), while the variable 'y' has the Gödel number 13. Substitute in this formula for the variable with Gödel number 13 (i.e., for 'y') the numeral for m. The result is the formula '$(\exists x)(x = sm)$', which says literally that there is a number x such that x is the immediate successor of m. This latter formula also has a Gödel number, which can be calculated quite easily. But instead of making the calculation, we can identify the number by an unambiguous meta-mathematical characterization: it is the Gödel number of the formula that is obtained from the formula with Gödel number m, by substituting for the variable with Gödel number 13 the numeral for m. This meta-mathematical characterization uniquely determines a definite number which is a certain arithmetical function of the numbers m and 13, where the function it-

self can be expressed within the formalized system.[23]

---

[23] This function is quite complex. Just how complex becomes evident if we try to formulate it in greater detail. Let us attempt such a formulation, without carrying it to the bitter end. It was shown on page 73 that m, the Gödel number of '$(\exists x)(x = sy)$', is

$$2^8 \times 3^4 \times 5^{11} \times 7^9 \times 11^8 \times 13^{11} \times 17^5 \times 19^7 \times 23^{13} \times 29^9.$$

To find the Gödel number of '$(\exists x)(x = sm)$' (the formula obtained from the preceding one by substituting for the variable 'y' in the latter the numeral for m) we proceed as follows: This formula contains the numeral 'm', which is a *defined* sign, and, in accordance with the content of footnote 18, m must be replaced by its defining equivalent. When this is done, we obtain the formula:

$$(\exists x)(x = \text{ssssss} \ldots \text{s}0)$$

where the letter 's' occurs $m + 1$ times. This formula contains only the elementary signs belonging to the fundamental vocabulary, so that its Gödel number can be calculated. To do this, we first obtain the series of Gödel numbers associated with the elementary signs of the formula:

$$8, 4, 11, 9, 8, 11, 5, 7, 7, 7, \ldots 7, 6, 9$$

in which the number 7 occurs $m + 1$ times. We next take the product of the first $m + 10$ primes in order of magnitude, each prime being raised to a power equal to the Gödel number of the corresponding elementary sign. Let us refer to this number as r, so that

$$r = 2^8 \times 3^4 \times 5^{11} \times 7^9 \times 11^8 \times 13^{11} \times 17^5 \times 19^7 \times 23^7 \times 29^7 \times 31^7 \times \ldots \times p^9_{m+10}$$

where p_{m+10} is the $(m + 10)$th prime in order of magnitude.

Now compare the two Gödel numbers m and r. m contains a prime factor *raised to the power 13*; r contains all the prime factors of m and many others besides, but *none of them are*

The number can therefore be designated *within* the calculus. This designation will be written as 'sub (m, 13, m)', the purpose of this form being to recall the meta-mathematical characterization which it represents, viz., 'the Gödel number of the formula obtained from the formula with Gödel number m, by substituting for the variable with the Gödel number 13 the numeral for m'. We can now drop the example and generalize. The reader will see readily that the expression 'sub (y, 13, y)' is the mirror image *within* the formalized arithmetical calculus of the meta-mathematical characterization: 'the Gödel number of the formula that is obtained from the formula with Gödel number y, by substituting for the variable with Gödel number 13 the numeral for y'. He will also note that when a definite numeral is substituted for 'y' in 'sub (y, 13, y)'—for example, the numeral for m or the numeral for two hundred forty three million—the resulting expression designates a definite integer which is the Gödel number of a certain formula.[24]

---

raised to the power 13. The number r can thus be obtained from the number m, by replacing the prime factor in m which is raised to the power 13 with other primes raised to some power different from 13. To state exactly and in full detail how r is related to m is not possible without introducing a great deal of additional notational apparatus; this is done in Gödel's original paper. But enough has been said to indicate that the number r is a definite arithmetical function of m and 13.

[24] Several questions may occur to the reader that need to be answered. It may be asked why, in the meta-mathematical characterization just mentioned, we say that it is "the *numeral* for y" which is to be substituted for a certain variable, rather than "the *number* y." The answer depends on the distinction, already discussed, between mathematics and meta-mathematics, and calls for a brief elucidation of the difference between numbers and numerals. A *numeral* is a *sign*, a linguistic expression, something which one can write down, erase, copy, and so on. A *number*, on the other hand, is something which a numeral *names* or *designates*, and which cannot literally be written down, erased, copied, and so on. Thus, we say that 10 is the *number* of our fingers, and, in making this statement, we are attributing a certain "property" to the class of our fingers; but it would evidently be absurd to say that this property is a numeral. Again, the number 10 is named by the Arabic numeral '10', as well as by the Roman letter 'X'; these names are different, though they name the same number. In short, when we make a substitution for a numerical variable (which is a letter or sign) we are putting one sign in place of another sign. We cannot literally substitute a number for a sign, because a number is a property of classes (and is sometimes said to be a concept), not something we can put on paper. It follows that, in substituting for a numerical variable, we can substitute only a numeral (or some other numerical expression, such as 's0' or '7 + 5'), and not a number. This explains why, in the above meta-mathematical characterization, we state that we are substituting for the variable the *numeral* for (the number) y, rather than the *number* y itself.

The reader may wonder what number is designated by 'sub (y, 13, y)' if the formula whose Gödel number is y does not happen to contain the variable with Gödel number 13—that is, if the formula does not contain the variable 'y'. Thus, sub (243,000,000, 13, 243,000,000) is the Gödel number of the formula obtained from the formula with Gödel number 243,000,000 by substituting for the variable 'y' the numeral '243,000,000'. But if the reader consults Table 4, he will find that 243,000,000 is the Gödel number of the formula '0 = 0', which does not contain the variable 'y'. What, then, is the formula that is obtained from '$0 = 0$' by substituting for the variable 'y' the numeral for the number 243,000,000? The simple answer is that, since '$0 = 0$' does not contain this variable, no substitution can be made—or, what amounts to the same thing, that the formula obtained from '$0 = 0$' is this *very same* formula. Accordingly, the number designated by 'sub (243,000,000, 13, 243,000,000)' is 243,000,000.

The reader may also be puzzled as to whether 'sub (y, 13, y)' is a *formula* within the arithmetical system in the sense that, for example, '$(\exists x)(x = sy)$', '$0 = 0$', and 'Dem (x, z)' are formulas. The answer is no, for the following reason. The expression '$0 = 0$' is called a formula, because it asserts a relation between two numbers and is thus capable of having truth or falsity significantly attributed to it. Similarly, when definite numerals are substituted for the variables in 'Dem (x, z)', this expression formulates a relation between two numbers, and so becomes a statement that is either true or false. The same holds for '$(\exists x)(x = sy)$'. On the other hand, even when a definite numeral is substituted for 'y' in 'sub (y, 13, y)', the resulting expression does not *assert* anything and therefore cannot be true or false. It merely *designates* or *names* a number, by describing it as a certain *function* of other numbers. The difference between a *formula* (which is in effect a statement about numbers, and so is either true or false) and a *name-function* (which is in effect a name that identifies a number, and so is neither true nor false) may be clarified by some further illustrations. '$5 = 3$' is a formula which, though false, declares that the two numbers 5 and 3 are equal; '$5^2 = 4^2 + 3^2$' is also a formula which asserts that a definite relation holds between the three numbers 5, 4, and 3; and, more generally, '$y = f(x)$' is a formula which asserts that a certain relation holds between the unspecified numbers x and y. On the other hand, '$2 + 3$' expresses a function of the two numbers 2 and 3, and so names a certain number (in fact, the number 5); it is not a formula, for it clearly would be nonsensical to ask whether '$2 + 3$' is true or false. '$(7 \times 5) + 8$' expresses another function of the three numbers 5, 7, and 8, and designates the number 43. And, more generally, '$f(x)$' expresses a function of x, and identifies a certain number when a definite numeral is substituted for 'x' and when a definite meaning is given to the function-sign 'f'. In short, while 'Dem (x, z)' is a formula because it has the *form of a statement* about numbers, 'sub (y, 13, y)' is not a formula because it has only the *form of a name* for numbers.

C *The heart of Gödel's argument*

At last we are equipped to follow in outline Gödel's main argument. We shall begin by enumerating the steps in a general way, so that the reader can get a bird's-eye view of the sequence.

Gödel showed (i) how to construct an arithmetical formula G that represents the meta-mathematical statement: 'The formula G is not demonstrable'. This formula G thus ostensibly says of *itself* that it is not demonstrable. Up to a point, G is constructed analogously to the Richard Paradox. In that Paradox, the expression 'Richardian' is associated with a certain number n, and the sentence 'n is Richardian' is constructed. In Gödel's argument, the formula G is also associated with a certain number h, and is so constructed that it corresponds to the statement: 'The formula with the associated number h is not demonstrable'. But (ii) Gödel also showed that G is demonstrable if, and only if, its formal negation $\sim$ G is demonstrable. This step in the argument is again analogous to a step in the Richard Paradox, in which it is proved that n is Richardian if, and only if, n is not

Richardian. However, if a formula and its own negation are both formally demonstrable, the arithmetical calculus is not consistent. Accordingly, if the calculus is consistent, neither G nor ~ G is formally derivable from the axioms of arithmetic. Therefore, if arithmetic is consistent, G is a formally undecidable formula. Gödel then proved (iii) that, though G is not formally demonstrable, it nevertheless is a *true* arithmetical formula. It is true in the sense that it asserts that every integer possesses a certain arithmetical property, which can be exactly defined and is exhibited by whatever integer is examined. (iv) Since G is both true and formally undecidable, the axioms of arithmetic are *incomplete*. In other words, we cannot deduce all arithmetical truths from the axioms. Moreover, Gödel established that arithmetic is *essentially* incomplete: even if additional axioms were assumed so that the true formula G could be formally derived from the augmented set, another true but formally undecidable formula could be constructed. (v) Next, Gödel described how to construct an arithmetical formula A that represents the meta-mathematical statement: 'Arithmetic is consistent'; and he proved that the formula 'A ⊃ G' is formally demonstrable. Finally, he showed that the formula A is not demonstrable. From this it follows that the consistency of arithmetic cannot be established by an argument that can be represented in the formal arithmetical calculus.

Now, to give the substance of the argument more fully:

(i) The formula '~ Dem (*x*, *z*)' has already been identified. It represents within formalized arithmetic the meta-mathematical statement: 'The sequence of formulas with the Gödel number *x* is not a proof for the formula with the Gödel number *z*'. The prefix '(*x*)' is now introduced into the Dem formula. This prefix performs the same function in the formalized system as does the English phrase 'For every *x*'. On attaching this prefix, we have a new formula: '(*x*) ~ Dem (*x*, *z*)', which represents within arithmetic the meta-mathematical statement: 'For every *x*, the sequence of formulas with Gödel number *x* is not a proof for the formula with Gödel number *z*'. The new formula is therefore the formal paraphrase (strictly speaking, it is the unique representative), within the calculus, of the meta-mathematical statement: 'The formula with Gödel number *z* is not demonstrable'—or, to put it another way, 'No proof can be adduced for the formula with Gödel number *z*'.

What Gödel showed is that a certain special case of this formula is not formally demonstrable. To construct this special case, begin with the formula displayed as line (1):

(1) (*x*) ~ Dem (*x*, sub (*y*, 13, *y*))

This formula belongs to the arithmetical calculus, but it represents a meta-mathematical statement. The question is, which one? The reader should first recall that the expression 'sub (*y*, 13, *y*)' designates a number. This number is the Gödel number of the formula obtained from the formula with Gödel number *y*, by substituting for the variable with Gödel number 13 the numeral for y.[25] It will then be evident that the formula of line (1) represents the meta-mathematical statement: 'The formula with Gödel number sub (y, 13, y) is not demonstrable'.[26]

---

[25] It is of utmost importance to recognize that 'sub (y, 13, y)', though it is an expression in formalized arithmetic, is not a formula but rather a name-function for identifying a *number* (see explanatory footnote 24). The number so identified, however, is the Gödel number of a formula—of the formula obtained from the formula with Gödel number y, by substituting for the variable 'y' the numeral for y.

[26] This statement can be expanded still further to read: 'The formula [whose Gödel number is the number of the formula] obtained from the formula with Gödel number y, by substituting for the variable with Gödel number 13 the numeral for y, is not demonstrable'.

The reader may be puzzled by the fact that, in the meta-mathematical statement 'The formula with Gödel number sub (y, 13, y) is not demonstrable', the expression 'sub (y, 13, y)' does not appear within quotation marks, although it has been repeatedly stated in the text that 'sub (y, 13, y)' is an *expression*. The point involved hinges once more on the distinction between using an expression to talk about what the expression designates (in which case the expression is not placed within quotation marks) and talking about the expression itself (in which case we must use a name for the expression and, in conformity with the convention for constructing such names, must place the expression within quotation marks). An example will help. '7 + 5' is an expression which designates a number; on the other hand, 7 + 5 is a number, and not an expression. Similarly, 'sub (243,000,000, 13, 243,000,000)' is an expression which designates the Gödel number of a formula (see Table 4); but sub (243,000,000, 13, 243,000,000) is the Gödel number of a formula, and is not an expression.

But, since the formula of line (1) belongs to the arithmetical calculus, it has a Gödel number that can actually be calculated. Suppose the number to be n. We now substitute for the variable with Gödel number 13 (i.e., for the variable 'y') in the formula of line (1) the numeral for n. A new formula is then obtained, which we shall call 'G' (after Gödel) and display under that label:

$$\text{(G)} \qquad (x) \sim \text{Dem}\ (x, \text{sub}\ (n, 13, n))$$

Formula G is the special case we promised to construct.

Now, this formula occurs within the arithmetical calculus, and therefore must have a Gödel number. What is the number? A little reflection shows that it is sub (n, 13, n). To grasp this, we must recall that sub (n, 13, n) is the Gödel number of the formula that is obtained from the formula with Gödel number n by substituting for the variable with Gödel number 13 (i.e., for the variable 'y') the numeral for n. But the formula G has been obtained from the formula with Gödel number n (i.e., from the formula displayed on line (1)) by substituting for the variable 'y' occurring in it the numeral for n. Hence the Gödel number of G is in fact sub (n, 13, n).

But we must also remember that the formula G is the mirror image *within* the arithmetical calculus of the meta-mathematical statement: 'The formula with Gödel number sub (n, 13, n) is not demonstrable'. It follows that the *arithmetical formula* '(x) ~ Dem (x, sub (n, 13, n))' *represents* in the calculus the *meta-mathematical statement:* 'The formula '(x) ~ Dem (x, sub (n, 13, n))' is not demonstrable'. In a sense, therefore, this arithmetical formula G can be construed as asserting of itself that it is not demonstrable.

(ii) We come to the next step, the proof that G is not formally demonstrable. Gödel's demonstration resembles the development of the Richard Paradox, but stays clear of its fallacious reasoning.[27] The argument is relatively unencumbered. It proceeds by showing that *if* the formula G were demonstrable then its formal

[27] It may be useful to make explicit the resemblance as well as the dissimilarity of the present argument to that used in the Richard Paradox. The main point to observe is that the formula G is not identical with the meta-mathematical statement with which it is associated, but only *represents* (or mirrors) the latter within the arithmetical calculus. In the Richard Paradox (as explained on p. 63 above) the number n is the number associated with a certain *meta-mathematical* expression. In the Gödel construction, the number n is associated with a certain *arithmetical formula* belonging to the formal calculus, though this arithmetical formula in fact represents a meta-mathematical statement. (The formula represents this statement, because the meta-mathematics of arithmetic has been mapped onto arithmetic.) In developing the Richard Paradox, the question is asked whether the number n possesses the *meta-mathematical* property of being Richardian. In the Gödel construction, the question asked is whether the number sub (n, 13, n) possesses a certain *arithmetical* property—namely, the arithmetical property expressed by the formula '(x) ~ Dem (x, z)'. There is therefore no confusion in the Gödel construction between statements *within* arithmetic and statements *about* arithmetic, such as occurs in the Richard Paradox.

contradictory (namely, the formula '~ (x) ~ Dem (x, sub (n, 13, n))') would also be demonstrable; and, conversely, that *if* the formal contradictory of G were demonstrable then G itself would also be demonstrable. Thus we have: G is demonstrable if, and only if, ~ G is demonstrable.[28] But as we noted earlier, if a

[28] This is not what Gödel actually proved; and the statement in the text, an adaptation of a theorem obtained by J. Barkley Rosser in 1936, is used for the sake of simplicity in exposition. What Gödel actually showed is that if G is demonstrable then ~ G is demonstrable (so that arithmetic is then inconsistent); and if ~ G is demonstrable then arithmetic is ω-inconsistent. What is ω-inconsistency? Let 'P' be some arithmetical predicate. Then arithmetic would be ω-inconsistent if it were possible to demonstrate both the formula '(∃x)P(x)' (i.e., 'There is at least one number that has the property P') and also each of the infinite set of formulas '~ P(0)', '~ P(1)', '~ P(2)', etc. (i.e., '0 does not have the property P', '1 does not have the property P', '2 does not have the property P', and so on). A little reflection shows that if a calculus is inconsistent then it is also ω-inconsistent; but the converse does not necessarily hold: a system may be ω-inconsistent without being inconsistent. For a system to be inconsistent, both '(∃x)P(x)' and '(x) ~ P(x)' must be demonstrable. However, although if a system is ω-inconsistent both '(∃x) P(x)' and each of the infinite set of formulas '~ P(0)', '~ P(1)', '~ P(2)', etc., are demonstrable, the formula '(x) ~ P(x)' may nevertheless not be demonstrable, so that the system is not inconsistent.

We outline the first part of Gödel's argument that if G is demonstrable then ~ G is demonstrable. Suppose the formula G were demonstrable. Then there must be a sequence of formulas within arithmetic that constitutes a proof for G. Let the Gödel number of this proof be k. Accordingly, the arithmetical relation designated by 'Dem (x, z)' must hold between formula and its formal negation can both be derived from a set of axioms, the axioms are not consistent. Whence, if the axioms of the formalized system of arithmetic are consistent, neither the formula G nor its negation is demonstrable. In short, if the axioms are consistent, G is formally *undecidable*—in the precise technical sense that neither G nor its contradictory can be formally deduced from the axioms.

(iii) This conclusion may not appear at first sight to be of capital importance. Why is it so remarkable, it may be asked, that a formula can be constructed within arithmetic which is undecidable? There is a surprise in store which illuminates the profound implications of this result. For, although the formula G is undecidable if the axioms of the system are consistent, it

---

k, the Gödel number of the proof, and sub (*n*, 13, *n*), the Gödel number of G, which is to say that 'Dem (*k*, sub (*n*, 13, *n*))' must be a true arithmetical formula. However, it can be shown that this arithmetical relation is of such type that, if it holds between a definite pair of numbers, the formula that expresses this fact is demonstrable. Consequently, the formula 'Dem (*k*, sub (*n*, 13, *n*))' is not only true, but also formally demonstrable; that is, the formula is a *theorem*. But, with the help of the Transformation Rules in elementary logic, we can immediately derive from this theorem the formula '∼ (*x*) ∼ Dem (*x*, sub (*n*, 13, *n*))'. We have therefore shown that if the formula G is demonstrable its formal negation is demonstrable. It follows that if the formal system is consistent the formula G is not demonstrable.

A somewhat analogous but more complicated argument is required to show that if ∼ G is demonstrable then G is also demonstrable. We shall not attempt to outline it.

can nevertheless be shown by *meta-mathematical* reasoning that G *is true*. That is, it can be shown that G formulates a complex but definite numerical property which necessarily holds of all integers—just as the formula '(*x*) ∼ (*x* + 3 = 2)' (which, when it is interpreted in the usual way, says that no cardinal number, when added to 3, yields a sum equal to 2) expresses another, likewise necessary (though much simpler) property of all integers. The reasoning that validates the truth of the undecidable formula G is straightforward. First, on the assumption that arithmetic is consistent, the meta-mathematical statement 'The formula '(*x*) ∼ Dem (*x*, sub (*n*, 13, *n*))' is not demonstrable' has been proven true. Second, this statement is represented within arithmetic by the very formula mentioned in the statement. Third, we recall that meta-mathematical statements have been mapped onto the arithmetical formalism in such a way that true meta-mathematical statements correspond to true arithmetical formulas. (Indeed, the setting up of such a correspondence is the *raison d'être* of the mapping; as, for example, in analytic geometry where, by virtue of this process, true geometric statements always correspond to true algebraic statements.) It follows that the formula G, which corresponds to a true meta-mathematical statement, must be true. It should be noted, however, that we have established an arithmetical truth, not by deducing it formally from the axioms of arithmetic, but by a meta-mathematical argument.

(iv) We now remind the reader of the notion of "completeness" introduced in the discussion of the sentential calculus. It was explained that the axioms of a deductive system are "complete" if every true statement that can be expressed in the system is formally deducible from the axioms. If this is not the case, that is, if not every true statement expressible in the system is deducible, the axioms are "incomplete." But, since we have just established that G is a true formula of arithmetic not formally deducible within it, it follows that the axioms of arithmetic are incomplete—on the hypothesis, of course, that they are consistent. Moreover, they are *essentially* incomplete: even if G were added as a further axiom, the augmented set would still not suffice to yield formally *all* arithmetical truths. For, if the initial axioms were augmented in the suggested manner, another true but undecidable arithmetical formula could be constructed in the enlarged system; such a formula could be constructed merely by repeating in the new system the procedure used originally for specifying a true but undecidable formula in the initial system. This remarkable conclusion holds, no matter how often the initial system is enlarged. We are thus compelled to recognize a fundamental limitation in the power of the axiomatic method. Against previous assumptions, the vast continent of arithmetical truth cannot be brought into systematic order by laying down once for all a set of axioms from which *every* true arithmetical statement can be formally derived.

(v) We come to the coda of Gödel's amazing intel-

lectual symphony. The steps have been traced by which he grounded the meta-mathematical statement: 'If arithmetic is consistent, it is incomplete'. But it can also be shown that this conditional statement *taken as a whole* is represented by a *demonstrable* formula within formalized arithmetic.

This crucial formula can be easily constructed. As we explained in Section V, the meta-mathematical statement 'Arithmetic is consistent' is equivalent to the statement 'There is at least one formula of arithmetic that is not demonstrable'. The latter is represented in the formal calculus by the following formula, which we shall call 'A':

(A) $$(\exists y)(x) \sim \text{Dem}\,(x, y)$$

In words, this says: 'There is at least one number y such that, for every number x, x does not stand in the relation Dem to y'. Interpreted meta-mathematically, the formula asserts: 'There is at least one formula of arithmetic for which no sequence of formulas constitutes a proof'. The formula A therefore represents the antecedent clause of the meta-mathematical statement 'If arithmetic is consistent, it is incomplete'. On the other hand, the consequent clause in this statement—namely, 'It [arithmetic] is incomplete'—follows directly from 'There is a true arithmetical statement that is not formally demonstrable in arithmetic'; and the latter, as the reader will recognize, is represented in the arithmetical calculus by an old friend, the formula G. Accordingly, the conditional meta-mathematical statement 'If arithmetic is consistent, it is incomplete' is represented by the formula:

$$(\exists y)(x) \sim \text{Dem}\,(x, y) \supset (x) \sim \text{Dem}\,(x, \text{sub}\,(n, 13, n))$$

which, for the sake of brevity, can be symbolized by '$A \supset G$'. (This formula can be proved formally demonstrable, but we shall not in these pages undertake the task.)

We now show that the formula A is not demonstrable. For suppose it were. Then, since $A \supset G$ is demonstrable, by use of the Rule of Detachment the formula G would be demonstrable. But, unless the calculus is inconsistent, G is formally undecidable, that is, not demonstrable. Thus if arithmetic is consistent, the formula A is not demonstrable.

What does this signify? The formula A represents the meta-mathematical statement 'Arithmetic is consistent'. If, therefore, this statement could be established by any argument that can be mapped onto a sequence of formulas which constitutes a proof in the arithmetical calculus, the formula A would itself be demonstrable. But this, as we have just seen, is impossible, if arithmetic is consistent. The grand final step is before us: we must conclude that if arithmetic is consistent its consistency cannot be established by any meta-mathematical reasoning that can be represented within the formalism of arithmetic!

This imposing result of Gödel's analysis should not be misunderstood: it does *not* exclude a meta-mathematical proof of the consistency of arithmetic. What it excludes is a proof of consistency that can be mirrored

by the formal deductions of arithmetic.[29] Meta-mathematical proofs of the consistency of arithmetic have, in fact, been constructed, notably by Gerhard Gentzen, a member of the Hilbert school, in 1936, and by others since then.[30] These proofs are of great logical significance, among other reasons because they propose new forms of meta-mathematical constructions, and because they thereby help make clear how the class of rules of inference needs to be enlarged if the consistency of arithmetic is to be established. But these proofs cannot be represented within the arithmetical calculus; and, since they are not finitistic, they do not achieve the proclaimed objectives of Hilbert's original program.

[29] The reader may be helped on this point by the reminder that, similarly, the proof that it is impossible to trisect an arbitrary angle with compass and straight-edge does *not* mean that an angle cannot be trisected by any means whatever. On the contrary, an arbitrary angle can be trisected if, for example, in addition to the use of compass and straight-edge, one is permitted to employ a fixed distance marked on the straight-edge.

[30] Gentzen's proof depends on arranging all the demonstrations of arithmetic in a linear order according to their degree of "simplicity." The arrangement turns out to have a pattern that is of a certain "transfinite ordinal" type. (The theory of transfinite ordinal numbers was created by the German mathematician Georg Cantor in the nineteenth century.) The proof of consistency is obtained by applying to this linear order a rule of inference called "the principle of transfinite induction." Gentzen's argument cannot be mapped onto the formalism of arithmetic. Moreover, although most students do not question the cogency of the proof, it is not finitistic in the sense of Hilbert's original stipulations for an absolute proof of consistency.

VIII

Concluding Reflections

The import of Gödel's conclusions is far-reaching, though it has not yet been fully fathomed. These conclusions show that the prospect of finding for every deductive system (and, in particular, for a system in which the whole of arithmetic can be expressed) an absolute proof of consistency that satisfies the finitistic requirements of Hilbert's proposal, though not logically impossible, is most unlikely.[31] They show also that there is an endless number of true arithmetical statements which cannot be formally deduced from any given set of axioms by a closed set of rules of inference. It follows that an axiomatic approach to number the-

[31] The possibility of constructing a finitistic absolute proof of consistency for arithmetic is not excluded by Gödel's results. Gödel showed that no such proof is possible that can be represented within arithmetic. His argument does not eliminate the possibility of strictly finitistic proofs that cannot be represented within arithmetic. But no one today appears to have a clear idea of what a finitistic proof would be like that is *not* capable of formulation within arithmetic.

ory, for example, cannot exhaust the domain of arithmetical truth. It follows, also, that what we understand by the process of mathematical proof does not coincide with the exploitation of a formalized axiomatic method. A formalized axiomatic procedure is based on an initially determined and fixed set of axioms and transformation rules. As Gödel's own arguments show, no antecedent limits can be placed on the inventiveness of mathematicians in devising new rules of proof. Consequently, no final account can be given of the precise logical form of valid mathematical demonstrations. In the light of these circumstances, whether an all-inclusive definition of mathematical or logical truth can be devised, and whether, as Gödel himself appears to believe, only a thoroughgoing philosophical "realism" of the ancient Platonic type can supply an adequate definition, are problems still under debate and too difficult for further consideration here.[32]

[32] Platonic realism takes the view that mathematics does not create or invent its "objects," but discovers them as Columbus discovered America. Now, if this is true, the objects must in some sense "exist" prior to their discovery. According to Platonic doctrine, the objects of mathematical study are not found in the spatio-temporal order. They are disembodied eternal Forms or Archetypes, which dwell in a distinctive realm accessible only to the intellect. On this view, the triangular or circular shapes of physical bodies that can be perceived by the senses are not the proper objects of mathematics. These shapes are merely imperfect embodiments of an indivisible "perfect" Triangle or "perfect" Circle, which is uncreated, is never fully manifested by material things, and can

Gödel's conclusions bear on the question whether a calculating machine can be constructed that would match the human brain in mathematical intelligence. Today's calculating machines have a fixed set of directives built into them; these directives correspond to the fixed rules of inference of formalized axiomatic procedure. The machines thus supply answers to problems by operating in a step-by-step manner, each step being controlled by the built-in directives. But, as Gödel showed in his incompleteness theorem, there are innumerable problems in elementary number theory that fall outside the scope of a fixed axiomatic method, and that such engines are incapable of answering, however intricate and ingenious their built-in mechanisms may be and however rapid their operations. Given a definite problem, a machine of this type might be built for solving it; but no one such machine can be built for solving every problem. The human brain may, to be sure, have built-in limitations of its own, and there may be mathematical problems it is incapable of solving. But, even so, the brain appears to embody a structure of rules of operation which is far more powerful

be grasped solely by the exploring mind of the mathematician. Gödel appears to hold a similar view when he says, "Classes and concepts may . . . be conceived as real objects . . . existing independently of our definitions and constructions. It seems to me that the assumption of such objects is quite as legitimate as the assumption of physical bodies and there is quite as much reason to believe in their existence" (Kurt Gödel, "Russell's Mathematical Logic," in *The Philosophy of Bertrand Russell* (ed. Paul A. Schilpp, Evanston and Chicago, 1944), p. 137).

than the structure of currently conceived artificial machines. There is no immediate prospect of replacing the human mind by robots.

Gödel's proof should not be construed as an invitation to despair or as an excuse for mystery-mongering. The discovery that there are arithmetical truths which cannot be demonstrated formally does not mean that there are truths which are forever incapable of becoming known, or that a "mystic" intuition (radically different in kind and authority from what is generally operative in intellectual advances) must replace cogent proof. It does not mean, as a recent writer claims, that there are "ineluctable limits to human reason." It does mean that the resources of the human intellect have not been, and cannot be, fully formalized, and that new principles of demonstration forever await invention and discovery. We have seen that mathematical propositions which cannot be established by formal deduction from a given set of axioms may, nevertheless, be established by "informal" meta-mathematical reasoning. It would be irresponsible to claim that these formally indemonstrable truths established by meta-mathematical arguments are based on nothing better than bare appeals to intuition.

Nor do the inherent limitations of calculating machines imply that we cannot hope to explain living matter and human reason in physical and chemical terms. The possibility of such explanations is neither precluded nor affirmed by Gödel's incompleteness theorem. The theorem does indicate that the structure and power of the human mind are far more complex and subtle than any non-living machine yet envisaged. Gödel's own work is a remarkable example of such complexity and subtlety. It is an occasion, not for dejection, but for a renewed appreciation of the powers of creative reason.

Appendix

Notes

1. (page 11) It was not until 1899 that the arithmetic of cardinal numbers was axiomatized, by the Italian mathematician Giuseppe Peano. His axioms are five in number. They are formulated with the help of three undefined terms, acquaintance with the latter being assumed. The terms are: '*number*', '*zero*', and '*immediate successor of*'. Peano's axioms can be stated as follows:

> 1. Zero is a number.
> 2. The immediate successor of a number is a number.
> 3. Zero is not the immediate successor of a number.
> 4. No two numbers have the same immediate successor.
> 5. Any property belonging to zero, and also to the immediate successor of every number that has the property, belongs to all numbers.

The last axiom formulates what is often called the "principle of mathematical induction."