GÖDEL'S PROOF

Although little known, it is a landmark of 20th-century thought. The proof brought to light certain astonishing limitations which have always been inherent in mathematics and mathematical logic

by Ernest Nagel and James R. Newman

In 1931 a young mathematician of 25 named Kurt Gödel published in a German scientific periodical a paper which was read only by a few mathematicians. It bore the forbidding title: "On Formally Undecidable Propositions of *Principia Mathematica* and Related Systems." It dealt with a subject that has never attracted more than a small group of investigators, and its reasoning was so novel and complex that it was unintelligible even to most mathematicians. But Gödel's paper has become a landmark of science in the 20th century. As "Gödel's proof," its general conclusions have become known to many scientists, and appreciated to be of revolutionary importance. Gödel's achievement has been recognized by many honors; not long after his paper appeared the young man was invited from Vienna to join the Institute for Advanced Study at Princeton, and he has been a permanent member of the Institute since 1938. When Harvard University awarded him an honorary degree in 1952, the citation described his proof as one of the most important advances in logic in modern times.

Gödel attacked a central problem in the foundations of mathematics. The axiomatic method invented by the Greeks has always been regarded as the strongest foundation for erecting systems of mathematical thinking. This method, as every student of logic knows, consists in assuming certain propositions or axioms (*e.g.*, if equals be added to equals, the wholes are equal) and deriving other propositions or theorems from the axioms. Until recent times the only branch of mathematics that was considered by most students to be established on sound axiomatic foundations was geometry. But within the past two centuries powerful and rigorous systems of axioms have been developed for other branches of mathematics, including the familiar arithmetic of whole numbers. Mathematicians came to hope and believe that the whole realm of mathematical reasoning could be brought into order by way of the axiomatic method.

Gödel's paper put an end to this hope. He confronted mathematicians with proof that the axiomatic method has certain inherent limitations which rule out any possibility that even the ordinary arithmetic of whole numbers can ever be fully systematized by its means. What is more, his proofs brought the astounding and melancholy revelation that it is impossible to establish the logical consistency of any complex deductive system except by assuming principles of reasoning whose own internal consistency is as open to question as that of the system itself.

Gödel's paper was not, however, altogether negative. It introduced into the foundations of mathematics a new technique of analysis which is comparable in fertility with René Descartes's historic introduction of the algebraic method into geometry. Gödel's work initiated whole new branches of study in mathematical logic. It provoked a reappraisal of mathematical philosophies, and indeed of philosophies of knowledge in general.

His epoch-making paper is still not widely known, and its detailed demonstrations are too complex to be followed by a nonmathematician, but the main outlines of his argument and conclusions can be understood. This article will recount the background of the problem and the substance of Gödel's findings.

The New Mathematics

The 19th century witnessed a tremendous surge forward in mathematical research. Many fundamental problems that had long resisted solution were solved; new areas of mathematical study were created; foundations were newly built or rebuilt for various branches of the discipline. The most revolutionary development was the construction of new geometries by replacing certain of Euclid's axioms with different ones. In particular the modification of Euclid's parallel axiom led to immensely fruitful results [see "The Straight Line," by Morris Kline; SCIENTIFIC AMERICAN, March]. It was this successful departure that stimulated the development of an axiomatic basis for other branches of mathematics which had been cultivated in a more or less intuitive manner. One important conclusion that emerged from this critical examination of the foundations of mathematics was that the traditional conception of mathematics as the "science of quantity" was inadequate and misleading. For it became evident that mathematics was most essentially concerned with drawing necessary conclusions from a given set of axioms (or postulates). It was thus recognized to be much more "abstract" and "formal" than had been traditionally supposed: more "abstract" because mathematical statements can be construed to be about anything whatsoever, not merely about some inherently circumscribed set of objects or traits of objects; more "formal" because the validity of a mathematical demonstration 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 postulates' descriptive terms plays no essential role in the process of deriving theorems. The question that confronts a 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 only whether the alleged conclusions are in fact the necessary logical consequences of the initial assumptions. This approach recalls Bertrand Russell's famous epigram: Pure mathematics is the subject in which we do not know what we are talking about, nor 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. As mathematics became more abstract, men's minds were emancipated from habitual connotations of language and could construct novel systems of postulates. Formalization led in fact to a great variety of systems of considerable mathematical interest and value. Some of these systems, it must be admitted, did not lend themselves to interpretations as obviously intuitive ("common sense") as those of Euclidean geometry or arithmetic, but this fact caused no alarm. Intuition, for one thing, is an elastic faculty. Our children will have no difficulty in accepting as intuitively obvious the paradoxes of relativity, just as we do not boggle at ideas which were regarded as wholly unintuitive a couple of generations ago. Moreover intuition, as we all know, is not a safe guide: it cannot be used safely as a criterion of either truth or fruitfulness in scientific explorations.

However, the increased abstractness of mathematics also raised a more serious problem. When a set of axioms is taken to be about a definite and familiar domain of objects, it is usually possible to ascertain whether the axioms are indeed true of these objects, and if they are true, they must also be mutually consistent. But the abstract non-Euclidean axioms appeared to be plainly false as descriptions of space, and, for that matter, doubtfully true of anything. Thus the problem of establishing the internal consistency of non-Euclidean systems was formidable. In Riemannian geometry, for example, the famous parallel postulate of Euclid is replaced by the assumption that through a given point outside a line *no* parallel to the line can be drawn in the same plane. Now suppose the question: Is the Riemannian set of postulates consistent? They are apparently not true of the ordinary space of our experience. How then is their consistency to be tested? How can one prove they will not lead to contradictory theorems?

A general method for solving this problem was proposed. The underlying idea was to find a "model" for the postulates so that each postulate was converted into a true statement about the model. The procedure goes something like this. Let us take the word "class" to signify a collection of distinguishable elements, or "members." (For example, the class of prime numbers less than 10 is a collection consisting of 2, 3, 5 and 7 as members.) Suppose now we consider two purely abstract classes, K and L, concerning which these postulates are given:

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 little set we can derive, by using customary rules of inference, certain theorems. For example, it can be shown that K contains just three members. But is the set a consistent one, so that mutually contradictory theorems can never be derived from it? This is where we invoke the help of a model, or interpretation, of the classes. Let K be the vertices of a triangle, and L its sides. Each of the five abstract postulates is then converted into a true statement: *e.g.*, the first postulate asserts that any two of the vertices are contained on just one side. In this way the set is proved to be consistent.

At first thought such a procedure may seem to suffice to establish the consistency of an abstract system such as plane Riemannian geometry. We may adopt a model embodying the Riemannian postulates in which the expression "plane" signifies 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 can then be converted into a theorem of Euclid. For example, on this interpretation the Riemannian parallel postulate reads as follows: 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.

Unhappily this method is vulnerable to a serious objection; namely, that it attempts to solve a problem in one domain merely by shifting the problem to another (or, to put it another way, we invoke Euclid to demonstrate the consistency of a system which subverts Eu-

All gentlemen are polite.
No bankers are polite.
No gentlemen are bankers.

$$g \subset p$$
$$b \subset \bar{p}$$
$$\therefore g \subset \bar{b}$$

$$g\bar{p} = 0$$
$$bp = 0$$
$$gb = 0$$

SYMBOLIC LOGIC was invented in the middle of the 19th century by the English mathematician George Boole. In this illustration a syllogism is translated into his notation in two different ways. In the upper group of formulas, the symbol $\subset$ means "is contained in." Thus $g \subset p$ says that the class of gentlemen is included in the class of polite persons. In the equations below two letters together mean the class of things having both characteristics. For example, *bp* means the class of individuals who are bankers and polite. The second equation in the group says that this class has no members. A line above a letter means "not." (Not-*p*, for example, means impolite.)

clid). Riemannian geometry is proved to be consistent only if Euclidean geometry is consistent. Query, then: Is Euclidean geometry consistent? If we attempt to answer this question by invoking yet another model, we are no closer to our goal. In short, any proof obtained by this method will be only a "relative" proof of consistency, not an absolute proof.

So long as we can interpret a system by a model containing only a finite number of elements, we have no great difficulty in proving the consistency of its postulates. For example, the triangle model which we used to test the K and L class postulates is finite, and accordingly it is comparatively simple to determine by actual inspection whether the postulates are "true" and hence consistent. Unfortunately most of the postulate systems that constitute the foundations of important branches of mathematics cannot be mirrored in finite models; they can be satisfied only by non-finite ones. In a well-known set of axioms for elementary arithmetic one of the axioms asserts that every integer in the sequence of whole numbers has an immediate successor which differs from any preceding integer. Obviously any model used to test the set of postulates must mirror the infinity of elements postulated by this axiom. It follows that the truth (and so the consistency) of the set cannot be established by inspection and enumeration. Apparently we have reached an impasse.

Russell's Paradox

It may be tempting to suggest at this point that we can be sure that a set of postulates is consistent, *i.e.*, free from contradictions, if the basic notions employed are transparently "clear" and "certain." But the history of thought has not dealt kindly with the doctrine of intuitive knowledge implicit in this suggestion. In certain areas of mathematical research 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 called "antinomies") have emerged, for example, in the theory of infinite numbers developed by Georg Cantor in the 19th century. His theory was built on the elementary and seemingly "clear" concept of class. Since modern systems in other branches of mathematics, particularly elementary arithmetic, have been built on the foundation of the theory of classes, it is pertinent to ask whether they, too, are not infected with contradictions.

In point of fact, Bertrand Russell constructed a contradiction within the framework of elementary logic itself. It is precisely analogous to the contradiction first developed in the Cantorian theory of infinite classes. Russell's antinomy can be stated as follows: All classes apparently may be divided into two groups: those which do not contain themselves as members, and those which do. An example of the first 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 the second is the class of all thinkable concepts, for the class of all thinkable concepts is itself a thinkable concept, and is therefore a member of itself. We shall call the first type of class "normal," and the second type "non-normal." Now let N stand for the class of all normal classes. We ask whether N itself is a normal class. If so, it is a member of itself. But in that case N is non-normal, because by definition a class which contains itself is non-normal. Yet if N is non-normal and thus a member of itself, it must be normal, because by definition all the members of N are normal. In short, N is normal if and only if N is non-normal. 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. Non-finite models by their very nature involve the use of possibly inconsistent sets of postulates. Thus it became clear that, although the model method for establishing the consistency of axioms is an invaluable mathematical tool, that method does not supply a final answer to the problem it was designed to resolve.

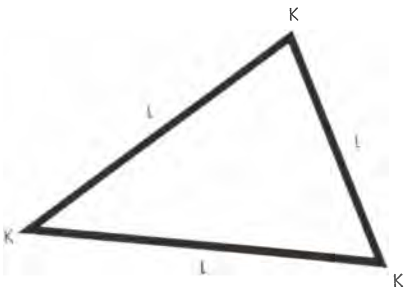
Hilbert's Meta-Mathematics

The eminent German mathematician David Hilbert then adopted the opposite approach of eschewing models and draining mathematics of any meaning whatever. In Hilbert's complete formalization, mathematical expressions are regarded simply as empty signs. The postulates and theorems constructed from the system of signs (called a calculus) are simply sequences of meaningless marks which are combined in strict agreement with explicitly stated rules. The derivation of theorems from postulates can be viewed as simply the transformation of one set of such sequences, or "strings," into another set of "strings," in accordance with precise rules of operation. In this manner Hilbert hoped to eliminate the danger of using any unavowed principles of reasoning.

Formalization is a difficult and tricky business, but it serves a valuable purpose. It reveals logical relations in naked clarity, as does a cut-away working model of a machine. One is able to see the structural patterns of various "strings" of signs: how they hang together, how they are combined, how they nest in one another, and so on. A page covered with the "meaningless" marks of such a formalized mathematics does not *assert* anything—it is simply an abstract design or a mosaic possessing a certain structure. But configurations of such a system can be described, and statements can be made about their various relations to one another. One mav say that a "string" is pretty, or that it resembles another "string," or that one "string" appears to be made up of three others, and so on. Such statements will evidently be meaningful.

Now it is plain that any meaningful statements about a meaningless system do not themselves belong to that system. Hilbert assigned them to a separate realm which he called "meta-mathematics." Meta-mathematical statements are statements *about* the signs and expressions of a formalized mathematical system: about the kinds and arrangements of such signs when they are combined to form longer strings of marks called "formulas," or about the relations between formulas which may obtain as a consequence of the rules of manipulation that have been specified for them.

A few examples will illustrate Hilbert's distinction between mathematics (a system of meaningless expressions) and meta-mathematics (statements about mathematics). Consider the arithmetical expression $2+3=5$. This expression belongs to mathematics and is constructed entirely out of elementary arithmetical signs. Now we may make a statement about the displayed expression, *viz.:* "'$2+3=5$' is an arithmetical formula." The statement does not express an arithmetical fact: it belongs to meta-mathematics, because it characterizes the string of arithmetical signs. Similarly the expression $x=x$ belongs to mathematics, but the statement "'x' is a variable" belongs to meta-mathematics. We may also make the following meta-mathematical statement: "The formula '$0=0$' is derivable from the formula '$x=x$' by substituting the numeral '0' for the variable 'x'." This statement specifies in what manner one arithmetical formula can be obtained from another formula, and thereby describes how the two formulas are related to each other. Again, we may make the meta-mathematical statement: "'$0\neq 0$' is not a theorem." It says that the formula in question is not derivable from the axioms of arithmetic, or in other words, that a certain relation does not hold between the specified formulas of the system. Finally, the following statement also belongs to meta-mathematics: "Arithmetic is consistent" (*i.e.*, it is not possible to derive from the axioms of arithmetic both the formula $0=0$ and also the formula $0\neq 0$).



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.

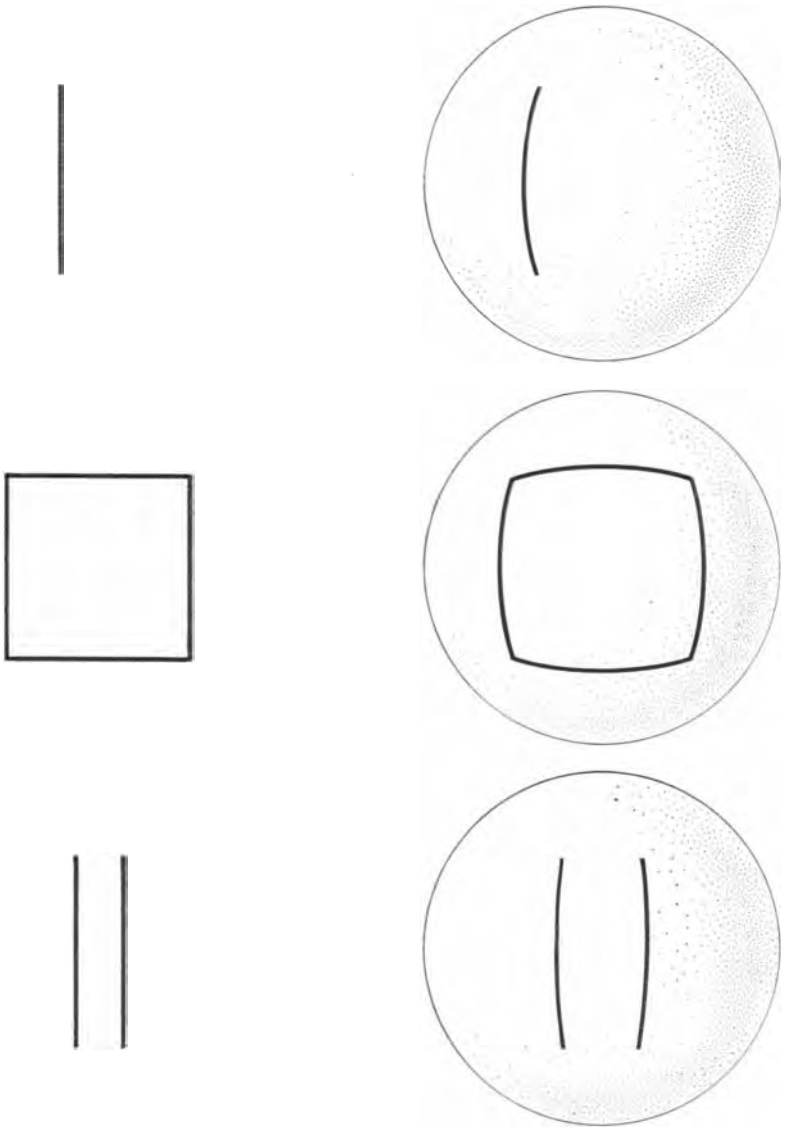
Upon this foundation—separation of meta-mathematical descriptions from mathematics itself—Hilbert attempted to build a method of "absolute" proof of the internal consistency of mathematical systems. Specifically, he sought to develop a theory of proof which would yield demonstrations of consistency by an analysis of the purely structural features of expressions in completely formalized (or "uninterpreted") calculi. Such an analysis consists exclusively of noting the kinds and arrangements of signs in formulas and determining whether a given combination of signs can be obtained from others in accordance with the explicitly stated rules of operation. An absolute proof of the consistency of arithmetic, if one could be constructed, would consist in showing by meta-mathematical procedures of a "finitistic" (non-infinite) character that two "contradictory" formulas, such as $(0=0)$ and its negation, cannot both be derived from the axioms or initial formulas by valid rules of inference.

It may be useful, by way of illustration, to compare meta-mathematics as a theory of proof with the theory of chess. Chess is played with 32 pieces of specified design on a square board containing 64 square subdivisions, where the pieces may be moved in accordance with fixed rules. Neither the pieces, nor the squares, nor the positions of the pieces on the board signify anything *outside* the game. In this sense the pieces and their configurations on the board are "meaningless." Thus the game is analogous to a formalized mathematical calculus. The pieces and the squares of the board correspond to the elementary signs of the calculus; the initial positions of the pieces correspond to the axioms or initial formulas of the calculus; their subsequent positions correspond to formulas derived from the axioms (*i.e.*, to the theorems), and the rules of the game correspond to the rules of inference for the calculus. Now, though configurations of pieces on the board are "meaningless," statements about these configurations, like meta-mathematical statements about mathematical formulas, are quite meaningful. A "meta-chess" statement may assert that there are 20 possible opening moves for White, or that, given a certain configuration of pieces on the board with White to move, Black is mate in three moves. Moreover, one can prove general "meta-chess" theorems on the basis of the finite number of permissible configurations on the board. The meta-chess theorem about the number of possible opening moves for White can be established in this way, and so can the meta-chess theorem that if White has only two Knights, it is impossible for White to mate Black. These and other "meta-chess" theorems can, in other words, be proved by finitistic methods of reasoning, consisting in the examination of each of a finite number of configurations that can occur under stated conditions. The aim of Hilbert's theory of proof, similarly, was to demonstrate by such finitistic methods the impossibility of deriving certain contradictory formulas in a calculus.

The Principia

It was Hilbert's approach, coupled with the formalization of logic itself in the famous *Principia Mathematica* by Alfred North Whitehead and Bertrand Russell, that led to the crisis to which Gödel supplied a final answer.

The grand object of *Principia*, published in 1910, was to demonstrate that mathematics is only a chapter of logic. But it made two contributions which are of particular interest to us here. First, following up work by the 19th-century pioneer George Boole, it supplied a system of symbols which permitted all statements of pure mathematics to be codified in a standard manner [see "Symbolic Logic," by John E. Pfeiffer; SCIENTIFIC AMERICAN, December, 1950]. Secondly, it stated in explicit form most of the rules of formal logic that are employed in mathematical proofs. Thus *Principia* provided an essential instrument for investigating the entire system of arithmetic as a system of "meaningless" marks which could be operated upon in accordance with explicitly stated rules.



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

We turn now to the formalization of a small portion of *Principia*, namely, the elementary logic of propositions. The task is to convert this fragment into a "meaningless" calculus of uninterpreted signs and to demonstrate a method of proving that the calculus is free from contradictions.

Four steps are involved. First we must specify the complete "vocabulary" of signs to be employed in the calculus. Second, we state the "formation rules" (the rules of "grammar") which indicate the combinations of signs permissible as formulas (or "sentences"). Third, we specify the "transformation rules," which tell how formulas may be derived from others. Finally, we select certain formulas as axioms which serve as foundations for the entire system. The "theorems" of the system are all the formulas, including the axioms, that can be derived from the axioms by applying the transformation rules. A "proof" consists of a finite sequence of legitimate formulas, each of which is either an axiom or is derivable from preceding formulas in the sequence by the transformation rules.

The vocabulary for the elementary logic of propositions (often also called the "sentential calculus") is extremely simple. The "sentential" variables (which correspond to sentences or statements) are certain letters: p, q, r and so on. Then there are several connectives: $\sim$, which stands for "not"; $\vee$, which stands for "or"; $\supset$, which stands for "if . . . then," and $\cdot$, which stands for "and." Parenthesis marks are used as signs of punctuation.

Each sentential variable counts as a formula, and the signs may be combined according to the formation rules to form other formulas: *e.g.*, $p \supset q$. If a given sentence $(p \supset q)$ is a formula, so is its negation $\sim (p \supset q)$. If two sentences, S_1 and S_2, are formulas, so is the combination $(S_1) \vee (S_2)$. Similar conventions apply to the other connectives.

For transformations there are just two rules. One, the rule of substitution, says that if a sentence containing sentential variables has been assumed, any formulas may be substituted everywhere for these variables, so that the new sentence will count as a logical consequence of the original one. For example, having accepted $p \supset p$ (if p, then p), we can

1	$(p \vee p) \supset$ If either *p* or *p*,	If either Henry VIII was a boor or Henry VIII was a boor, then Henry VIII was a boor.
2	$p \supset (p \vee q)$ If *p*,	If psychoanalysis is valid then either psychoanalysis is valid or headache powders are better.
3	$(p \vee q) \supset (q \vee p)$ If either *p* or *q*, then either *q* or *p*	If either Immanuel Kant was punctual or Hollywood is sinful then either Hollywood is sinful or Immanuel Kant was punctual.
4	$(p \supset q) \supset [(r \vee p) \supset$ If *p* implies *q*, then (either *r* or *p*)	If ducks waddle implies that $\sqrt{2}$ is a number then (either Churchill drinks brandy or ducks waddle) implies brandy or $\sqrt{2}$ is a number).

SENTENTIAL CALCULUS, or the elementary logic of propositions, is based on four axioms. The nonsense statements illustrate how general is the "meaning" of the symbols.

always substitute q for p, obtaining as a theorem the formula $q \supset q$; or we may substitute $(p \vee q)$ for p, obtaining $(p \vee q) \supset (p \vee q)$. The other rule, that of detachment, simply says that if the sentences S_1 and $S_1 \supset S_2$ are logically true, we may also accept as logically true the sentence S_2.

The calculus has four axioms, essentially those of *Principia*, which are given in the table at the top of this page, along with nonsensical English sentences to illustrate their independence of meaning. The clumsiness of the translations, especially in the case of the fourth axiom, will perhaps help the reader to realize the advantages of using a special symbolism.

Search for a Proof

Each of these axioms may seem "obvious" and trivial. Nevertheless it is possible to derive from them with the help of the stated transformation rules an indefinitely large class of theorems which are far from obvious or trivial. However, at this point we are interested not in deriving theorems from the axioms but in showing that this set of axioms is not contradictory. We wish to prove that, using the transformation rules, it is impossible to derive from the axioms any formula S (*i.e.*, any expression which would normally count as a sentence) together with its negation $\sim$ S.

Now it can be shown that $p \supset (\sim p \supset q)$ (if p, then if not-p then q) is a theorem in the calculus. Let us suppose, for the sake of demonstration, that a formula S and its contradictory $\sim$ S were both deducible from the axioms, and test the consequences by means of this theorem. By substituting S for p in the theorem, as permitted by the rule of substitution, we first obtain $S \supset (\sim S \supset q)$. From this, assuming S to be demonstrably true, we could next obtain, by the detachment rule, $\sim S \supset q$. Finally, if we assume $\sim$ S also is demonstrable, by the detachment rule we would get q. Since we can substitute any formula whatsoever for q, this means that any formula whatsoever would be deducible from the axioms. Thus if both S and its contradictory $\sim$ S were deducible from the axioms, then *any* formula would be deducible. We arrive, then, at the conclusion that if the calculus is not consistent (*i.e.*, if both S and $\sim$ S are deducible) any theorem can be derived from the axioms. Accordingly, to prove the consistency of the calculus, our task is reduced to finding at least one formula which cannot be derived from the axioms.

The way this is done is to employ meta-mathematical reasoning upon the system before us. The actual procedure is elegant. It consists in finding a characteristic of formulas which satisfies the three following conditions. (1) it is common to all four axioms; (2) it is "hereditary," that is, any formula derived from the axioms (*i.e.*, any theorem) must also have the property; (3) there must be at least one formula which does not have the characteristic and is therefore not a theorem. If we succeed in this threefold task, we shall have an absolute proof of the consistency of the axioms. If we can find an array of signs that conforms to the requirements of being a formula but does not possess the specified characteristic, this formula cannot be a theorem. In other words, the finding of a single formula which is not a theorem suffices to establish the consistency of the system.

Let us choose as a characteristic of the required kind the property of being a "tautology." In common parlance a tautology is usually considered to be a redundant statement such as: "John is the father of Charles and Charles is a son of John." But in logic a tautology is defined as a statement which excludes no logical possibilities—*e.g.*, "Either it is raining or it is not raining." Another way of putting this is to say that a tautology is "true in all possible worlds." We apply this definition to formulas in the system we are considering. A formula is said to be a tautology if it is invariably true regardless of whether its elementary constituents (p, q, r and so on) are true or false. Now all four of our axioms plainly possess the property of being tautologous. For example, the first axiom, $(p \vee p) \supset p$, is true regardless of whether p is assumed to be true or is assumed to be false. The axiom says, for instance:

CONNECTIVES AND ELEMENTARY SIGNS

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 next following number
(	8	punctuation mark
)	9	punctuation mark
,	10	punctuation mark

SENTENTIAL VARIABLES (EACH DESIGNATED BY A NUMBER GREATER THAN 10 AND DIVISIBLE BY 3)

VARIABLES	GÖDEL NUMBER	SAMPLE
p	12	Henry VIII was a boor.
q	15	Headache powders are better.
r	18	Ducks waddle.
etc.		

INDIVIDUAL VARIABLES (EACH DESIGNATED BY A NUMBER GREATER THAN 10 WHICH LEAVES A REMAINDER OF 1 WHEN DIVIDED BY 3)

VARIABLES	GÖDEL NUMBER	MEANING
x	13	a numerical variable
y	16	a numerical variable
z	19	a numerical variable
etc.		

PREDICATE VARIABLES (EACH DESIGNATED BY A NUMBER GREATER THAN 10 WHICH LEAVES A REMAINDER OF 2 WHEN DIVIDED BY 3)

VARIABLES	GÖDEL NUMBER	SAMPLE
P	14	Being a boor
Q	17	Being a headache powder
R	20	Being a duck
etc.		

ELEMENTARY GÖDEL NUMBERS are assigned to every symbol used in his system of symbolic logic in accordance with the orderly scheme which is illustrated in the table above.

"If either Mount Rainier is 20,000 feet high or Mount Rainier is 20,000 feet high, then Mount Rainier is 20,000 feet high." It makes no difference whether Mount Rainier is actually 20,000 feet high or not: the statement is still true in either case. A similar demonstration can be made for the other axioms.

Next it is possible to prove that the property of being a tautology is hereditary under the transformation rules, though we shall not turn aside to give the demonstration. It follows that every formula properly derived from the axioms (*i.e.*, every theorem) must be a tautology. Having performed these two steps, we are ready to look for a formula which does not possess the characteristic of being a tautology. We do not have to look very hard. For example, $p \vee q$ fits the requirements. Clearly it is not a tautology; it is the same as saying: "Either John is a philosopher or Charles reads SCIENTIFIC AMERICAN." This is patently not a truth of logic; it is not a sentence that is true irrespective of the truth or falsity of its elementary constituents. Thus $p \vee q$, though it purports to be a gosling, is in fact a duckling; it is a formula but it is not a theorem.

We have achieved our goal. We have found at least one formula which is not a theorem, therefore the axioms must be consistent.

Gödel's Answer

The sentential calculus is an example of a mathematical system for which the objectives of Hilbert's theory of proof are fully realized. But this calculus codifies only a fragment of formal logic. The question remains: Can a formalized system embracing the whole of arithmetic be proved consistent in the sense of Hilbert's program?

This was the conundrum that Gödel answered. His paper in 1931 showed that all such efforts to prove arithmetic to be free from contradictions are doomed to failure.

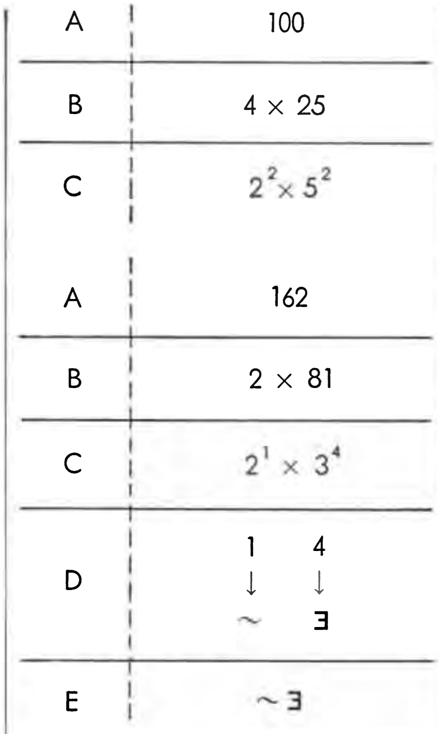
His main conclusions were twofold. In the first place, he showed that it is impossible to establish a meta-mathematical proof of the consistency of a system comprehensive enough to contain the whole of arithmetic—unless, indeed, this proof itself employs rules of inference much more powerful than the transformation rules used in deriving theorems within the system. In short, one dragon is slain only to create another.

Gödel's second main conclusion was even more surprising and revolutionary, for it made evident a fundamental limitation in the power of the axiomatic method itself. Gödel showed that *Principia*, or any other system within which arithmetic can be developed, is essentially incomplete. In other words, given *any* consistent set of arithmetical axioms, there are true arithmetical statements which are not derivable from the set. A classic illustration of a mathematical "theorem" which has thwarted all attempts at proof is that of Christian Goldbach, stating that every even number is the sum of two primes. No even number has ever been found which is not the sum of two primes, yet no one has succeeded in finding a proof that the rule applies without exception to all even numbers. In reply to Gödel it might be suggested that the set of arithmetical axioms could be modified or expanded to make "underivable" statements derivable. But Gödel showed that this approach promises no final cure. That is, even if any finite number of other axioms is added, there will always be further arithmetical truths which are not formally derivable.

How did Gödel prove his conclusions? His paper is difficult. A reader must master 46 preliminary definitions, together with several important preliminary theorems, before he gets to the main results. We shall take a much easier road; nevertheless we hope at least to offer glimpses of the argument.

Gödel Numbers

Gödel first devised a method of assigning a number as a label for each elementary sign, each formula and each proof in a formalized system. To the elementary signs he attached as "Gödel numbers" the integers from 1 to 10; to the variables he assigned numbers according to certain rules [*see table at left*]. To see how a number is given to a formula of the system, let us take this formula: $(\exists\, x)\ (x = Sy)$, which reads literally "there is an *x*, such that *x* is the immediate successor of *y*" and in effect says that every number has an immediate successor. The numbers associated with the formula's 10 successive signs are, respectively, 8, 4, 13, 9, 8, 13, 5, 7, 16, 9 [*see table*]. Now these numbers are to be used as exponents, or powers, of the first 10 prime numbers (*i.e.*, 2, 3, 5 and so on). The prime numbers, raised to these powers, are multiplied together. Thus we get the number $2^8 \times 3^4 \times 5^{13} \times 7^9 \times 11^8 \times 13^{13} \times 17^5 \times 19^7 \times 23^{16} \times 29^9$. The product is the Gödel number of the formula. In the same way every formula can be represented by a single unique number.



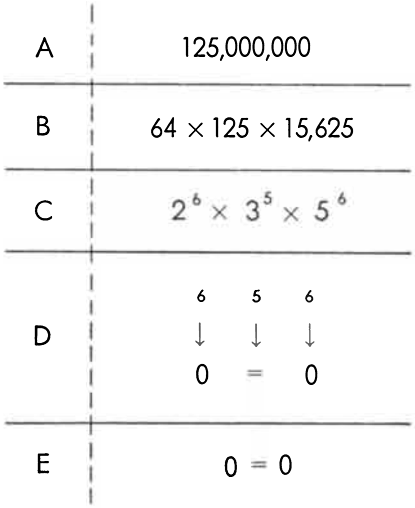
GÖDEL NUMBERS of formulas are constructed by raising the prime numbers, in sequence, to powers which are the Gödel numbers of the symbols involved. Thus 100 is not a Gödel number because its factors skip the prime number 3. On the other hand, 162 is the Gödel number for "there is not."

We can assign a number to a sequence of formulas, such as may occur in some proof, by a similar process. Let us say that we have a sequence of two formulas, the second derived from the first. For example, by substituting 0 for *y* in the formula given above, we derive $(\exists\, x)\ (x = S0)$, which says that 0 has an immediate successor. Now the first and second formulas are identified by Gödel numbers which we shall call *m* and *n*, respectively. To label this sequence, we use the Gödel numbers *m* and *n* as exponents and multiply the first two primes (2 and 3) raised to these powers. That is to say, the Gödel number that identifies the sequence is $2^m \times 3^n$. In like manner we can give a number to any sequence of formulas or any other expression in the system.

What has been done so far is to establish a method for completely arithmetizing a formal system. The method is essentially a set of directions for making a one-to-one correspondence between specific numbers and the various elements or combinations of elements of the system. Once an expression is given, it can be uniquely numbered. But more than that, we can retranslate any Gödel number into the expression it represents by factoring it into its component prime numbers, which can be done in only one way, as we know from a famous theorem of arithmetic [*see illustration below*]. In other words, we can take the number apart as if it were a machine, see how it was constructed and what went into it, and we can dissect an expression or a proof in the same way.

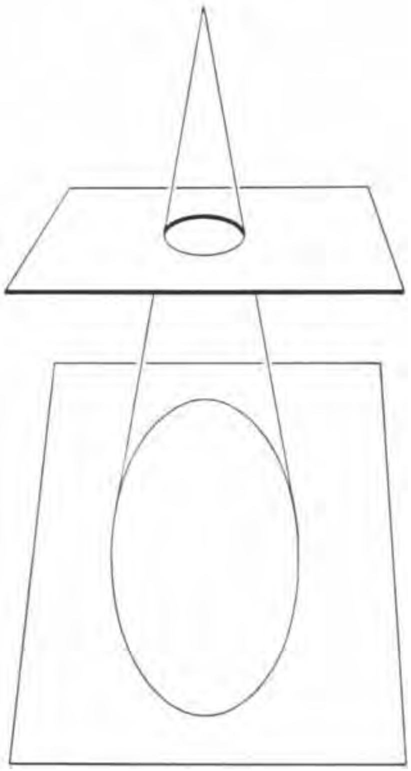
This leads to the next step. It occurred to Gödel that meta-mathematical statements can be translated into arithmetical terms by a process analogous to mapping. In geography the spatial relations between points on the spherical earth can be projected onto a flat map; in mathematical physics relations between the properties of electric currents can be mapped in terms of the flow of fluids; in mathematics itself relations in geometry can be translated into algebra. Gödel saw that if complicated meta-mathematical statements about a system could be translated into, or mirrored by, arithmetical statements within the system itself, an important gain would be achieved in clarity of expression and facility of analysis. Plainly it would be easier to deal with arithmetical counterparts of complex logical relations than with the logical relations themselves. To cite a trivial analogy: If customers in a supermarket are given tickets with numbers determining the order in which they are to be waited on, it is a simple matter to discover, merely by scrutinizing the numbers, how many persons have been served, how many are waiting, who precedes whom and by how many customers, and so on.

What Gödel aimed at was nothing less than the complete arithmetization of meta-mathematics. If each meta-mathe-



ARITHMETICAL FORMULA "zero equals zero" has the Gödel number 125 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.

matical statement could be uniquely represented in the formal system by a formula expressing a relation between numbers, questions of logical dependence between meta-mathematical statements could be explored by examining the corresponding relations between integers. Gödel did in fact succeed brilliantly in mapping the meta-mathematics of arithmetic upon arithmetic itself. We need cite only one illustration of how a meta-mathematical statement can be made to correspond to a formula in the formal arithmetical system. Let us take the formula $(p \vee p) \supset p$. We may make the meta-mathematical statement that the formula $(p \vee p)$ is the initial part of this formula. Now we can represent this meta-mathematical statement by an arithmetical formula which says in effect that the Gödel number of the initial part is a factor of the Gödel number of the complete formula. Evidently this is so, for the Gödel number of $(p \vee p)$ is $2^8 \times 3^{12} \times 5^2 \times 7^{12} \times 11^9$, while the Gödel number of $(p \vee p) \supset p$ is $2^8 \times 3^{12} \times 5^2 \times 7^{12} \times 11^9 \times 13^3 \times 17^{12}$.



MAPPING of objects from one realm onto another is illustrated above. Points in the upper, horizontal plane can be uniquely mapped onto the lower plane, which slants downward from back to front, by drawing lines from a single point through the points of the upper plane and extending them until they intersect the lower plane. Thus a circle in the upper plane maps as an ellipse in the lower. Gödel mapped statements about arithmetic as expressions in arithmetic.

The Undecidable Proposition

We have now arrived at the very heart of Gödel's analysis. He showed how to construct an arithmetical formula, whose Gödel number we shall suppose is h, which corresponds to the meta-mathematical statement, *viz.:* "The formula with Gödel number h is not demonstrable." In other words, this formula (call it G) in effect asserts its own indemonstrability, though it is a legitimate formula belonging to the formal system of arithmetic. Gödel then proceeded to examine the question whether G is or is not a demonstrable formula of arithmetic. He was able to show that G is demonstrable if, and only if, its negation, $\sim$ G, also is demonstrable. But if a formula and its negation are both derivable from a set of axioms, obviously the axioms are not consistent. It follows that if arithmetic is consistent, neither G nor its negation is demonstrable. That is to say, G is an undecidable formula of arithmetic. Now from this Gödel proved the indemonstrability of the proposition that arithmetic is consistent. It can be shown that a meta-mathematical statement of arithmetic's consistency corresponds to a certain arithmetical formula, A, and that the arithmetical formula $A \supset G$ (if A, then G) is demonstrable. Thus if A were demonstrable, G would be also. But we have just seen that G is not demonstrable. It follows that A is undecidable. In short, the consistency of arithmetic is undecidable by any meta-mathematical reasoning which can be represented within the formalism of arithmetic.

Gödel's analysis does not exclude a meta-mathematical demonstration of the consistency of arithmetic; indeed, such proofs have been constructed, notably by Gerhard Gentzen, a member of the Hilbert school. But these "proofs" are in a sense pointless, because they employ rules of inference whose own internal consistency is as much open to doubt as is the formal consistency of arithmetic itself. Gentzen's proof employs a rule of inference which in effect permits a formula to be derived from an infinite class of premises. And the employment of this non-finitistic meta-mathematical notion raises once more the difficulty which Hilbert's original program was intended to resolve.

There is another surprise coming. Although the formula G is undecidable, it can be shown by meta-mathematical reasoning that G is nevertheless a *true* arithmetical statement and expresses a property of the arithmetical integers. The argument for this conclusion is quite simple. We need recall only that Gödel mapped meta-mathematical statements upon arithmetical formulas in such a way that every true meta-mathematical statement corresponds to a true arithmetical formula. Now G corresponds to a meta-mathematical statement ("the formula with Gödel number h is not demonstrable") which, as we have seen, is true, unless arithmetic is inconsistent. It follows that G itself must be true. We have thus established an *arithmetical* truth by a *meta-mathematical* argument.

So we come to the finale of Gödel's amazing and profound intellectual symphony. Arithmetic is incomplete, in the transparent sense that there is at least one arithmetical truth which cannot be derived from the arithmetical axioms and yet can be established by a meta-mathematical argument outside the system. Moreover, arithmetic is *essentially* incomplete, for even if the true formula G were taken as an axiom and added to the original axioms, the augmented system would still not suffice to yield formally all the truths of arithmetic: we could still construct a true formula which would not be formally demonstrable within the system. And such would be the case no matter how often we repeated the process of adding axioms to the initial set.

This remarkable conclusion makes evident an inherent limitation in the axiomatic method. Contrary to previous assumptions, the vast "continent" of arithmetical truth cannot be brought into systematic order by way of specifying once for all a fixed set of axioms from which all true arithmetical statements would be formally derivable.

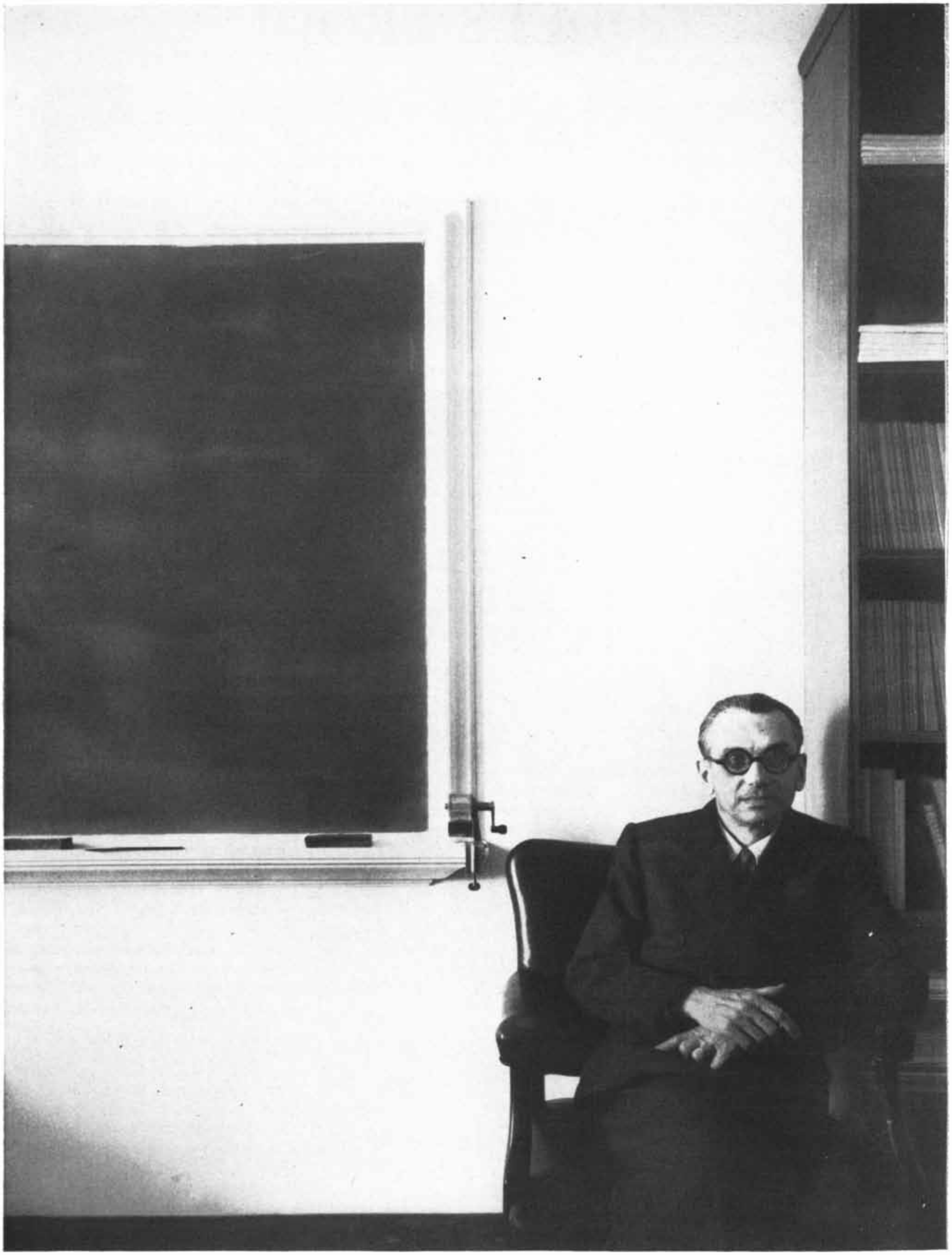
Men and Calculating Machines

The far-reaching import of Gödel's conclusions has not yet been fully fathomed. They show that the hope of finding an absolute proof of consistency for any deductive system expressing the whole of arithmetic cannot be realized, if such a proof must satisfy the finitistic requirements of Hilbert's original program. They also show that there is an endless number of true arithmetical statements which cannot be formally deduced from any specified set of axioms in accordance with a closed set of rules of inference. It follows that an axiomatic approach to the theory of numbers, for example, cannot exhaust the domain of arithmetic truth. Whether an all-inclusive general definition of mathematical or logical truth can be devised, and whether, as Gödel himself appears to believe, only a thoroughgoing Platonic realism can supply such a definition, are problems still under debate.

Gödel's conclusions have a bearing on the question whether a calculating machine can be constructed that would equal the human brain in mathematical reasoning. Present calculating machines have a fixed set of directives built into them, and they operate in a step-by-step manner. But in the light of Gödel's incompleteness theorem, there is an endless set of problems in elementary number theory for which such machines are inherently incapable of supplying answers, however complex their built-in mechanisms may be and however rapid their operations. The human brain may, to be sure, have built-in limitations of its own, and there may be mathematical problems which it is incapable of solving. But even so, the human brain appears to embody a structure of rules of operation which is far more powerful 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. 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 must replace cogent proof. 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.

Nor does the fact that it is impossible to construct a calculating machine equivalent to the human brain necessarily mean 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 discouragement but for a renewed appreciation of the powers of creative reason.



KURT GÖDEL was photographed in his office at the Institute for Advanced Study by Arnold Newman. Gödel was born in Czechoslovakia in 1906. He received his doctorate from the University of Vienna in 1930 and served on its faculty until he came to the U. S.