The Incompleteness Theorem

Martin Davis

In September 1930 in Königsberg, on the third day of a symposium devoted to the foundations of mathematics, the young Kurt Gödel launched his bombshell announcing his incompleteness theorem. At that time, there were three recognized "schools" on the foundations of mathematics: the logicism based on the work of Frege, Russell, and Whitehead that saw mathematics as simply part of logic, Brouwer's radical intuitionism, and Hilbert's proof theory (also called "formalism"). In fact two days earlier, lectures representing these schools had been delivered by Carnap, Heyting, and von Neumann respectively. Von Neumann may have been the only person in the room to have grasped the significance of what Gödel had done. He saw that the goals of Hilbert's proof theory had been shown to be simply unattainable. Logicism had also been dealt a death blow, but Carnap, who had known about Gödel's incompleteness theorem for over a week when he gave his address, seemed not to realize its significance.

Formalization of Mathematics

It was Gottlob Frege in his *Begriffsschrift* of 1879 who had shown how the logical reasoning used in mathematical proofs can be reduced to the combinatorial manipulation of symbols. By the 1920s foundational work had made it clear that the full expanse of classical mathematics could be encapsulated in such formal combinatorial systems. In these systems, a proposition of mathematics was represented by a string of symbols, and a proof, by a finite sequence of such strings. Since these systems were simple combinatorial objects, it seemed quite possible to apply mathematical methods to study their properties. Hilbert's program aimed to prove, by utterly unimpeachable methods, that these systems were consistent and complete: that they were safe from the catastrophic inconsistency, due to Russell's paradox, that had struck Frege's ambitious attempt to bridge the gap between the elements of formal logic and mathematics proper, and that with respect to some specified class of statements, each statement of the class could be either proved or refuted within the system. Gödel's incompleteness theorem did away with the second of these goals, and shortly thereafter Gödel was able to show that the first was likewise unachievable. Gödel's theorem had made it clear that no single formal system could be devised that would enable all mathematical truths, even those expressible in terms of basic operations on the natural numbers, to be provided with a formal proof.

Martin Davis is professor emeritus of mathematics and computer science, New York University, and is a Visiting Scholar in mathematics at the University of California, Berkeley. His email address is martin@eipye.com.

Gödel's Proof

Gödel proceeded to define a code by means of which each expression of a formal system would have its own natural number, what has come to be called its *Gödel number*, associated with it. Thus, expressions of the system that represent propositions about the natural numbers might be seen by someone privy to the code as also making assertions, incidentally as it were, about the system itself. Working with a particular formal system loosely based on that of Whitehead and Russell and exploiting this idea, Gödel showed how to construct a remarkable expression of the system we

may designate as U. To someone who didn't know the code, U would be seen as expressing a complicated and peculiar statement about the natural numbers. But to someone who could decipher it, U would be seen as also asserting that some statement expressible in the system is unprovable. Looking more closely, it would be found that the statement asserted to be unprovable is U itself. Thus we may say:

U asserts that it is unprovable.

Thus, if U were false, it would be provable, and hence, presumably, true. This contradiction shows that U is true, and hence, given what it asserts, unprovable. *There are true statements unprovable in the given system.*

Of course, this heuristic outline would have hardly been convincing. But Gödel carefully worked out the details leaving no doubt about the correctness of his conclusions.[1] Nevertheless, a whiff of paradox hung over the matter; it seemed hard to believe that a trick so close to puzzles usually offered for amusement could really be used to demonstrate something profound about mathematics.

Computability Theory Makes a Contribution

We write $N = \{0, 1, 2, \dots\}$ for the set of natural numbers. A function $f : N \to N$ is called *computable* if there is an algorithm that given an $x \in N$ will compute $f(x)$. Here the notion of algorithm is assumed to involve no restriction as to the amount of time or space required to complete a computation.[2] Finally a set $S \subseteq N$ is called *computable* if its characteristic function

$$C_S(x) = \begin{cases} 1 & \text{if } x \in S \\ 0 & \text{otherwise} \end{cases}$$

is computable.

The following is fundamental:

Theorem. *There is a computable function f whose range*

$$K = \{f(0), f(1), f(2), \dots\}$$

is not computable.[3]

Computability theory provides a perspective from which it can be seen that incompleteness is a pervasive fundamental property not dependent on a trifling trick. From this point of view the formal systems studied by logicians are simply computable functions that spew out theorems (more precisely, Gödel numbers of theorems). Such systems are usually given in terms of a set of axioms and rules of inference. One can then imagine an algorithm that begins with the axioms and proceeds by iteratively applying the rules of inference.

To obtain a form of the incompleteness theorem let us begin with the set K whose existence is given by the theorem above, and consider propositions of the form $n \notin K$ where n is a fixed natural number. We can suppose that, in a particular formal system these propositions are each represented by a corresponding string of symbols we may write as P_n. We need only assume that there is an algorithm for obtaining P_n given n.[4] Let us use the symbol $\mathcal{F}$ for some formal system, and write $\vdash_{\mathcal{F}} P_n$ to mean that P_n is provable in $\mathcal{F}$. We will say that $\mathcal{F}$ is *sound* if

Whenever $\vdash_{\mathcal{F}} P_n$ for a particular n,
it will also be the case that $n \notin K$.

Since P_n is intended to stand for the proposition $n \notin K$, soundness simply means that the provable statements are true.

Incompleteness Theorem. *Let $\mathcal{F}$ be a sound formal system. Then there is a number n_0 such that $n_0 \notin K$, but it is not the case that $\vdash_{\mathcal{F}} P_{n_0}$.*

Again, we have a true sentence that is not provable. Note that we only succeed in changing the value of the particular number n_0 as we attempt to create stronger and stronger formal systems that can prove more and more.

Proof of the Incompleteness Theorem. Suppose that there is no such n_0. Then we would have:

$\vdash_{\mathcal{F}} P_n$ for a particular n, if and only if $n \notin K$.

Recall that K is the range of the computable function f. Then the following would be an algorithm for computing $C_K(n)$ for a given value of n, contradicting the fact that K is not computable: Begin generating the theorems of $\mathcal{F}$ and at the same time begin computing the successive values $f(0), f(1), f(2), \dots$. If $n \in K$, then n will eventually show up in the list of values of f so $C_K(n) = 1$. Otherwise, P_n will eventually show up in the theorem list of $\mathcal{F}$ so that $C_K(n) = 0$. □

[1] *Detailed proofs can be found in a number of textbooks, for example [3]. In addition Gödel's clear and meticulous original exposition [8] still repays study.*

[2] *Computability theory has provided a number of precise characterizations to replace this heuristic explanation, and they have all been proved equivalent to one another.*

[3] *See for example [1]. Computability theory is also known as recursion theory and used to also be called recursive function theory. Computable functions are also called recursive. Sets that are the range of a computable function as well as the empty set are called recursively enumerable, or more recently, computably enumerable, or listable.*

[4] *In a traditional formal system, for a given number n, P_n will be obtained by replacing, in a certain specific formula, a symbol for a variable by a "numeral" representing the number n.*

A Diophantine Perspective

The following result, known variously as MRDP and as Matiyasevich's Theorem, enables it to be seen that the truths unprovable in specified formal systems can have a straightforward mathematical form.

Theorem. *If S is the range of a computable function, then there is a polynomial $p(a, x_1, \ldots, x_m)$ with integer coefficients such that the equation $p(a, x_1, \ldots, x_m) = 0$ has a solution in natural numbers $x_1, \ldots, x_m$ for a given value of a if and only if $a \in S$ (see [9, 2]).*

Applying this result to the case $S = K$, let us call the corresponding polynomial p_0. Now we can think of the expressions P_n as standing for the proposition that $p_0(n, x_1, \ldots, x_m) = 0$ has no solutions in natural numbers, and say that $\mathcal{F}$ is *Diophantine-sound* if $\vdash_{\mathcal{F}} P_n$ implies that the equation $p_0(n, x_1, \ldots, x_m) = 0$ does indeed fail to have solutions. Then, the incompleteness theorem of the previous section takes the form:

Diophantine Incompleteness Theorem. *Let $\mathcal{F}$ be Diophantine-sound. Then there is a number n_0 such that the equation $p_0(n_0, x_1, \ldots, x_m) = 0$ has no solutions in natural numbers although it is not the case that $\vdash_{\mathcal{F}} P_{n_0}$.*

It is worth remarking that the proof of MRDP is entirely constructive so the polynomial p_0 could be produced quite explicitly.

Two Formal Systems: PA and ZFC

What Frege showed is that the ordinary reasoning in proofs of mathematical theorems amounts to formal manipulations of the propositional connectives $\neg \ \rightarrow \ \vee \ \wedge$ together with the quantifiers $\forall \ \exists$. Manipulations of the propositional connectives amounts to carrying out the operations of Boolean algebra. The quantifiers get in the way of this, and careful rules are needed to justify removing and reinstating them. Once these rules are specified (which can be done in a number of equivalent ways), the way is open to set up formal systems encapsulating greater or lesser portions of mathematics. This involves supplying a vocabulary of symbols representing various constants, functions, and relations appropriate to the part of mathematics being formalized. Finally a list of axioms must be given: these are written using this vocabulary together with the symbols listed above corresponding to the operations of logical inference. A symbol for equality should also be available.[5]

For the system PA (for "Peano Arithmetic"), the vocabulary consists of symbols for the number 0, and for the successor, sum, and product functions on the natural numbers. The axioms are the familiar Peano postulates together with equations serving to implicitly define sum and product. The induction postulate, whose informal statement is that a set of natural numbers containing 0 and closed under successor must consist of all natural numbers, appears in a restricted form: it is stated only for sets *definable* in terms of the vocabulary.[6]

PA formalizes the elementary number theory of the textbooks as well as (via clever coding) substantial parts of elementary analysis. In contrast ZFC formalizes the full scope of modern set-theoretic mathematics including such things as general topology and transfinite arithmetic. The vocabulary can be extremely parsimonious consisting only of the symbol $\in$ for set membership. For our purposes it will be useful to be slightly less frugal, allowing as well symbols $\emptyset$ (the empty set), $\{\ldots\}$ (the set consisting of a single element), and $\cup$ (binary union). The axioms are those of Zermelo-Fraenkel together with the axiom of choice, and the resulting system is powerful enough to encapsulate the full scope of classical mathematics, and indeed, much more (see for example [4]).

We write $\vdash_{\mathrm{PA}}$ and $\vdash_{\mathrm{ZFC}}$ for provability in PA and ZFC, respectively. We will use $\mathcal{F}$ as a subscript to refer ambiguously to either of these systems. Also we write $\nvdash_{\mathcal{F}}$ to express non-provability in the corresponding systems.

In each of PA and ZFC, a simple notation is available for representing the natural numbers by sequences of strings we call *numerals*. We will write $\overline{n}$ for the numeral representing the natural number n. In PA this may be defined as follows using the letter s for successor and letting 0 be represented by its usual symbol:

$$\overline{0} = 0 \ ; \ \overline{n+1} = s\overline{n}$$

Following von Neumann, the numerals in ZFC can be defined as follows:

$$\overline{0} = \emptyset; \ \overline{n+1} = \overline{n} \cup \{\overline{n}\}$$

Now, for $\mathcal{F}$ standing for either PA or ZFC, associated with the polynomial p_0 of the previous section, there is a formula $\pi_{\mathcal{F}}(x_0, x_1, \ldots, x_m)$ such that for arbitrary natural numbers $a, a_1, \ldots, a_m$:

$$\text{if } p_0(a, a_1, \ldots, a_m) = 0 \text{ then } \vdash_{\mathcal{F}} \pi(\overline{a}, \overline{a_1}, \ldots, \overline{a_m})$$
$$\text{if } p_0(a, a_1, \ldots, a_m) \neq 0 \text{ then } \vdash_{\mathcal{F}} \neg\pi(\overline{a}, \overline{a_1}, \ldots, \overline{a_m})$$

For PA, this is almost a triviality because symbols for addition and multiplication are part of its vocabulary, and the axioms justify ordinary calculations. For ZFC, some circumlocution is needed, but

[5] *Equality can be thought of as the most "advanced" part of the underlying logic, or as the most fundamental mathematical relation.*

[6] *A fuller account of PA will be found in Feferman's article [5] in this issue of the* Notices. *Full details will be found in textbooks such as [3].*

the ordinary facts about addition and multiplication of natural numbers can still be replicated.

Incompleteness Theorem for PA and ZFC. *If $\mathcal{F}$ is consistent, then there is a natural number n_0 such that*

$$\nvdash_{\mathcal{F}} (\forall x_1)\ldots(\forall x_m)\neg\pi(\overline{n_0}, x_1, \ldots, x_m),$$

and

$$\nvdash_{\mathcal{F}} \neg(\forall x_1)\ldots(\forall x_m)\neg\pi(\overline{n_0}, x_1, \ldots, x_m).$$

Thus the *sentence* $(\forall x_1)\ldots(\forall x_m)\neg\pi(\overline{n_0}, x_1, \ldots, x_m)$ is *undecidable* in $\mathcal{F}$: neither it nor its negation is provable. However, what that sentence asserts, namely that the equation $p_0(n_0, x_1, \ldots, x_m) = 0$ has no solutions in natural numbers, is true. Moreover that truth is a consequence of its undecidability. For if $p_0(n_0, a_1, \ldots, a_m) = 0$ we would have $\vdash_{\mathcal{F}} \pi(\overline{n_0}, \overline{a_1}, \ldots, \overline{a_m})$ and using elementary logic we would obtain

$$\vdash_{\mathcal{F}} (\exists x_1)\ldots(\exists x_m)\pi(\overline{n_0}, x_1, \ldots, x_m)$$

from which we readily obtain

$$\vdash_{\mathcal{F}} \neg(\forall x_1)\ldots(\forall x_m)\neg\pi(\overline{n_0}, x_1, \ldots, x_m)$$

contradicting the claimed undecidability.

There has been much confusion about this situation. How is it that *we* can see that the proposition is true although a system as powerful as ZFC cannot? The answer is that ZFC can indeed see what we can, namely that *if* ZFC is consistent then the proposition is true but undecidable by its means. In fact, it was precisely by analyzing this situation that Gödel could conclude that systems like PA and ZFC cannot prove their own consistency, thereby shattering Hilbert's hopes.

The fact that ZFC is stronger than PA (in actual fact very much stronger) is exemplified by the following result:

Theorem. *If PA is consistent, then there is a natural number n_0 such that*

$$\nvdash_{\mathrm{PA}} (\forall x_1)\ldots(\forall x_m)\neg\pi(\overline{n_0}, x_1, \ldots, x_m),$$

and

$$\nvdash_{\mathrm{PA}} \neg(\forall x_1)\ldots(\forall x_m)\neg\pi(\overline{n_0}, x_1, \ldots, x_m),$$

but

$$\vdash_{\mathrm{ZFC}} (\forall x_1)\ldots(\forall x_m)\neg\pi(\overline{n_0}, x_1, \ldots, x_m).$$

So the undecidability in PA is decided in ZFC! But then ZFC has its own undecidability and with the very same formula π. Only the number n_0 changes. The values of n_0 for either system will be enormous since all the complexity of the algorithms for generating theorems these systems provide must be contained in those numbers.

We will refer to statements to the effect that some polynomial equation has no solutions in natural numbers as $\forall$-statements.[7] They all have the property just exhibited that their undecidability in a reasonable formal system implies their truth. It follows from the MRDP theorem that statements asserting that some computable property holds for all natural numbers are provably equivalent (for example in PA) to a $\forall$-statement. Many famous problems are thus seen to belong to this class, in particular, Fermat's last theorem, the Goldbach conjecture, the four-color theorem, and the Riemann Hypothesis (see [2]).

Beyond ZFC

At the same time that the ZFC axioms provide a foundation for mathematics, they also can be regarded as defining a class of mathematical structures. From this point of view they can be seen as providing "closure" under such operations as forming the set of all subsets of a given set or the union of all of its elements. In a normal situation of this kind it would be natural to find the least set closed under all of the operations called for by the axioms. Remarkably, as natural as such an object appears, its existence cannot be proved in ZFC. This is because if such existence could be proved, it would provide a model for the axioms and hence lead to a proof in ZFC of its own consistency. And this, Gödel had proved to be impossible. So systems like ZFC lead in a natural way to extensions, and in each such extension new $\forall$-propositions become provable. In fact there will be new values of n_0 for which the fact that the equation $p_0(n_0, x_1, \ldots, x_m) = 0$ has no solutions in natural numbers becomes provable. What remains unclear is whether some really mathematically significant $\forall$-propositions, perhaps like some of those mentioned at the end of the previous section, require means beyond ZFC for their proof. We conclude this article with a recent example announced by Harvey Friedman (see [7]) of a $\forall$-proposition that is unprovable in ZFC, but becomes provable with the aid of a so-called "large cardinal" axiom, an assumption of the existence of an infinite set of a size larger than any whose existence can be proved in ZFC.[8]

Friedman's example concerns finite directed graphs (no multiple edges allowed) whose vertices are finite sequences of integers. What is striking is that what looks like a harmless additional conclusion in a theorem provable in ZFC (and even in much weaker systems) results in a proposition that is unprovable in ZFC but becomes provable on the addition of a large cardinal axiom, an assumption of the existence of a set too large for that existence to be provable in ZFC.[9] Some preliminary definitions are needed. For a natural number n we write

[7] *Logicians call these Π^0_1 statements. As this notation suggests, they find their place in a hierarchy.*

[8] *The article by Juliet Floyd and Akihiro Kanamori [6] in this issue of the* Notices *contains some discussion of large cardinal axioms.*

[9] *Specifically, an axiom of the Mahlo type.*

$\hat{n}$ for the set $\{1, 2, \ldots, n\}$. So $\hat{n}^k$ is the set of all sequences of these numbers of length k. If $x, y \in \hat{n}^k$, we write $x * y$ for the element of $\hat{n}^{2k}$ obtained by concatenating x and y. We will work with directed graphs G whose vertex set $V(G)$ consists of elements of $\hat{n}^k$ for certain fixed n, k. For $x, y \in V(G)$ we write (x, y) for a possible edge proceeding from x to y. G is called an *upgraph* if for every edge (x, y) of G, we have $\max(x) < \max(y)$. We say that $u, v \in \hat{n}^\ell$ are *order equivalent* if for all $1 \leq i, j \leq \ell$, we have $u_i < u_j$ if and only if $v_i < v_j$. An upgraph G is called *order invarant* if whenever $x * y$ is order equivalent to $z * w$, we have (x, y) is an edge of G if and only if (z, w) is an edge of G. For $A \subseteq V(G)$ we write $GA = \{y \mid \exists x \in A$ say that (x, y) is an edge of $G\}$. A is called *independent* if no two elements of A are connected by an edge of G. Sets $B, C \subseteq V(G)$ are *G-isomorphic* if there is a bijection h from B to C such that for all $x, y \in B$, (x, y) is an edge in G if and only if (hx, hy) is also an edge in G. Finally, we call $x \in V(G)$ *two-powered* if each x_i is a member of the set $\{1, 2, 4, 8, \ldots\}$ of powers of 2. Now, we have:

Theorem. *For all $n, k, r \geq 1$ every order-invariant upgraph G on $\hat{n}^k$ has an independent set A such that if $B \subseteq V(G) - A$ and $|B| \leq r$, then B is G-isomorphic to a set $C \subseteq GA$ such that B and C have the same two-powered elements.*

This is provable not only in ZFC but also in PA and even in still weaker systems. However consider the following variant:

Proposition. *For all $n, k, r \geq 1$ every order-invariant upgraph G on $\hat{n}^k$ has an independent set A such that if $B \subseteq V(G) - A$ and $|B| \leq r$, then B is G-isomorphic to a set $C \subseteq GA$ such that B and C have the same two-powered elements, and furthermore, the particular number $2^{(4kr)^2} - 1$ doesn't occur in any element of C.*

Harvey Friedman has announced that this $\forall$-statement is not provable in ZFC but becomes provable on the addition of a large cardinal axiom.

Acknowledgments: I'm grateful to Solomon Feferman and to Allyn Jackson for their helpful comments on a previous version of this article.

References

[1] Nigel Cutland, *Computability: An Introduction to Recursive Function Theory,* Cambridge University Press, Cambridge, England, and New York, 1980.

[2] Martin Davis, Yuri Matijasevic, and Julia Robinson, "Hilbert's Tenth Problem: Diophantine Equations: Positive Aspects of a Negative Solution", *Proceedings of Symposia in Pure Mathematics,* vol. 28 (1976), pp. 323–378; reprinted in Feferman, Solomon, ed. *The Collected Works of Julia Robinson,* Amer. Math. Soc. 1996, pp. 269–378.

[3] Herbert Enderton, *A Mathematical Introduction to Logic,* Academic Press, New York, 1972.

[4] ____ , *Elements of Set Theory,* Academic Press, New York, 1977.

[5] Solomon Feferman, The Impact of the Incompleteness Theorems on Mathematics, *Notices*, April 2006.

[6] Juliet Floyd and Akihiro Kanamori, How Gödel Transformed Set Theory, *Notices*, April 2006.

[7] Harvey Friedman, "Π^0_1 Incompleteness: Finite Graph Theory 1." `http://www.math.ohio-state.edu/%7Efriedman/pdf/Pi01013006.pdf`.

[8] Kurt Gödel, "Über formal unentscheidbare Sätze der *Principia Mathematica* und verwandte Systeme I" with page-facing English translation, in Solomon Feferman, et al. (eds.), Kurt Gödel, *Collected Works, vol. I.,* Oxford, New York, 1986.

[9] Yuri Matiyasevich, *Hilbert's Tenth Problem,* MIT Press, Cambridge, Massachusetts, 1983.