

# GÖDEL'S THEOREM

## *Limits on Logic*

In the early 1930s, Kurt Gödel, a German mathematician, attempted to show that the predicate calculus (see Chapter 58) was “complete” — that one can obtain mechanically (in principle, at least) a proof of any true formula expressed in that calculus. His failure to do this was crowned by the discovery that the task was impossible: Certain formal systems, including arithmetic, were incomplete in this sense. The discovery staggered the world of mathematics.

As part of David Hilbert's turn-of-the-century program for mathematics, it had been expected that all mathematics, when it was suitably formalized in a system like the predicate calculus, would turn out to be complete. But Gödel discovered that not even arithmetic was complete. His now-famous theorem states that in any sound, consistent, formal system containing arithmetic there are true statements that cannot be proved — statements the truth of which we may know by other means but not by any formal, step-by-step decision process.

From Hilbert's time onward, a succession of researchers had attempted to formulate such decision processes in various ways. They had in common, however, the adoption of certain axioms and a variety of formalisms for manipulating the axioms by certain rules to obtain new statements that, given the truth of

the axioms, were true themselves. One such system was the theory of recursive functions, which Gödel helped to develop. Recursive functions amount to yet one more description of what it means to compute. (See Chapter 66.)

The heart of Gödel's method was to encode each possible statement in the predicate calculus, regarded as a language, by a special numerical code called its *Gödel number*. Briefly, the process has three steps.

Set up axioms for the predicate calculus along with a rule of inference by which one can get new formulas from old ones.

Set up axioms for standard arithmetic in the language of predicate calculus. Define a numbering for each formula or sequence of formulas in the resulting formal system.

By using easy-to-read implicative language, the axioms for the predicate calculus may be listed as follows:

1.  $\forall y_i(F \rightarrow (G \rightarrow F))$
2.  $\forall y_i((F \rightarrow (G \rightarrow H)) \rightarrow ((F \rightarrow G) \rightarrow (F \rightarrow H)))$
3.  $\forall y_i((\neg F \rightarrow \neg G) \rightarrow ((\neg F \rightarrow G) \rightarrow F))$
4.  $\forall y_i(\forall x(F \rightarrow G) \rightarrow (F \rightarrow \forall xG))$  provided  $F$  has no free occurrence of  $x$
5.  $\forall y_i((F \rightarrow G) \rightarrow (\forall y_i F \rightarrow \forall y_i G))$
6.  $\forall y_i(\forall x F(x) \rightarrow F(y))$  provided that  $y$  is not quantified when it is substituted

Once the notation used above is understood, the axioms will seem reasonably self-evident, the very thing upon which to base a theory of mathematical deduction.

Symbols such as  $F$ ,  $G$ , and  $H$  are understood to stand for "formulas";  $\forall y_i$  stands for an arbitrary string of variables such as  $y_1, y_2, \dots, y_k$ , all universally quantified; one reads such symbolism as "for all  $y_i$ " or "for all  $y_1, y_2, \dots, y_k$ ." The symbol  $\rightarrow$  stands for implication and  $\neg$  stands for negation.

The various formulas involved in the axioms above may or may not contain the variables which are quantified outside them, but in any case axiom 4 cannot be applied unless whenever variable  $x$  occurs in formula  $F$ , it must be quantified within  $F$ . Axiom 6 cannot be applied unless  $y$  is not quantified within  $F(y)$  (the formula one gets by replacing all free occurrences of  $x$  in  $F$  by  $y$ ).

The axioms themselves are now easily understood. For example, axiom 1 may be read, "For all possible values of their free variables, if  $F$  is true, then  $G \rightarrow F$ ." In other words, a true formula is implied by *any* formula. Axiom 2 says that implication distributes over itself, in effect. And axiom 3 says that if both  $\neg G$  and  $G$  are implied by  $\neg F$ , then  $\neg F$  cannot be true, in other words,  $F$  must be true.

As a final example, axiom 4 says, "For all possible values of their free variables (as always), if  $F \rightarrow G$  for all  $x$  and if  $x$  has no free occurrence in  $F$ , then  $F \rightarrow \forall xG$ ."

To the previous set of axioms must be added a rule of inference like the following:

If  $F$  and  $F \rightarrow G$ , then  $G$ .

One notes that this rule does not lie at the same level as the axioms, in a sense: It is intended that whenever we are making a deduction — essentially a chain of formulas — and notice formulas  $F$  and  $F \rightarrow G$  both occurring as earlier members of the chain, we may add  $G$  to the chain.

By the "standard arithmetic" is meant simply the *Peano postulates* for the natural numbers, namely,

1.  $\forall x \neg(0 = sx)$
2.  $\forall x, y (sx = sy) \rightarrow (x = y)$
3.  $\forall x x + 0 = x$
4.  $\forall x, y x + sy = s(x + y)$
5.  $\forall x, y x \times sy = x \times y + x$
6.  $\forall x x \times 0 = 0$

Here,  $s$  denotes the successor function which for each natural number  $x$  picks out its successor,  $x + 1$ . Thus, postulates 1 and 2 tell us, respectively, that

Zero is not the successor of any natural number.

If the successors are equal, then so are the numbers.

The remaining postulates are similarly easy to understand. However, all six postulates beg the question of what we mean by equality. This very meaning is embedded in three additional postulates:

7.  $\forall x x = x$
8.  $\forall x, y, z (x = y) \rightarrow ((x = z) \rightarrow (y = z))$
9.  $\forall x, y (x = y) \rightarrow (A(x, x) \rightarrow A(x, y))$

where  $A$  is any formula having two free variables.

Just as we added a special rule of inference to the axioms for the predicate calculus, so here we add a rule of induction:

$$(P(0) \& \forall x (P(x) \rightarrow P(sx))) \rightarrow \forall x P(x)$$

This formula simply encodes the well-known rule of induction: If a predicate  $P$  is true with 0 substituted in it *and* if whenever  $P$  is true of a number  $x$ ,  $P$  is also true of  $x$ 's successor, then  $P$  is true for all possible numbers  $x$ .

The 15 axioms and two rules listed above are, among them, powerful enough to give us a formal system for arithmetic in which so many ideas are expressible and truths provable that we are tempted a priori to imagine that any arithmetic truth is not only expressible in this system, but also provable there.

Having set up these axioms, Gödel went on to the third step in his proof, namely, to assign a unique number to every conceivable formula in the system just defined. He did this by assigning a natural number to each of the following basic symbols:

Symbol	Code Number	Symbol	Code Number
0	1	$x$	9
$s$	2	1	10
+	3	$\neg$	11
$\times$	4	&	12
=	5	$\exists$	13
(	6	$\forall$	14
)	7	$\rightarrow$	15
,	8		

Now, given any axiom or formula within our formal system, it is a straightforward matter to scan the formula from left to right and to replace each symbol in it by a prime number raised to the power of that symbol's code number. The primes used for this purpose are the consecutive primes 2, 3, 5, 7, 11, . . . . As an example of this procedure, axiom 4 of the standard arithmetic has the following Gödel number:

$$x_1 + sx_{11} = s(x_1 + x_{11})$$

$$2^9 \cdot 3^{10} \cdot 5^3 \cdot 7^2 \cdot 11^9 \cdot 13^{10} \cdot 17^{10} \cdot 19^5 \cdot 23^2 \cdot 29^6 \cdot 31^9 \cdot 37^{10} \cdot 41^3 \cdot 43^9 \cdot 47^{10} \cdot 53^{10} \cdot 59^7 \cdot$$

The number was obtained by scanning the given expression one symbol at a time and converting it to the appropriate prime power. Thus  $x$ , the first symbol, has code number 9; so 2, the first prime, is raised to the ninth power. The next symbol, 1, becomes  $3^{10}$  because 3 is the next prime and 10 is the code number for the symbol 1.

Note that the axiom has been altered to accommodate our use above of a special notation for variables, namely, to use a sort of unary code (consisting of

consecutive 1s) to yield a system of subscripts for the symbol  $x$ . In other words,  $x_1$  and  $x_{11}$  may be considered as perfectly general names like  $x$  and  $y$  in axiom 4 of the arithmetic.

As can be seen in the example above, Gödel numbers tend to be enormous. Nevertheless, they are computable, and given any integer whatsoever, it is possible to compute the expression it represents (if any) by finding all its prime factors and grouping these as powers in order of increasing primes.

We are now ready to enter the very heart of Gödel's theorem by considering the following predicate:

$$\text{Proof}(x,y,z)$$

Here,  $\text{Proof}(x,y,z)$  is a predicate having the following interpretation: " $x$  is the Gödel number of a proof  $X$  of a formula  $Y$  (with one free variable and Gödel number  $y$ ) which has the integer  $z$  substituted into it." The "proof  $X$ " referred to here may itself be considered a formula for the purpose of having a Gödel number assigned to it.

Note that the basic symbols to which code numbers were attached did not include the predicate symbol "Proof" or any other predicate symbol except equality ( $=$ ). Indeed, " $\text{Proof}(x,y,z)$ " is just *our own shorthand* for an immensely long expression with three free variables  $x$ ,  $y$ , and  $z$ —or, in Gödel's notation,  $x_1$ ,  $x_{11}$ , and  $x_{111}$ . This expression includes a number of procedures, including the following:

1. Given an integer, produce the string of which it is the Gödel number.
2. Given a string, check whether it is a formula.
3. Given a sequence of formulas, check whether it is a proof for the last formula in the sequence.

All these procedures are computable and, as Gödel showed, themselves reducible to formulas within the formal system defined above. Before we show how this predicate is used in Gödel's theorem, there is one small detail to clean up. Namely, we must state what a "formula" really is: The definition embedded in procedure 2 above would amount to an inductive definition of a properly formed arithmetic expression and the ways in which such expressions may be combined legally with the logical connectives  $\&$  and  $\rightarrow$  or quantified over by  $\exists$  and  $\forall$ . For example,  $\exists x_1(X_{11}(x) = X_{111}(x))$  is a formula but  $1X)\exists x_{11}(\neg =)$  is not.

We are now ready to consider a very special use of the predicate under consideration. Suppose that the formula  $Y$  is fed its own Gödel number and that

we deny the existence of a proof within the formal system of the resulting formula:

$$\neg \exists x \text{Proof}(x, y, y)$$

In words, the formula  $\text{Proof}(x, y, y)$  says, “ $x$  is the Gödel number of a proof of the formula obtained by substituting its own Gödel number  $y$  for its one free variable.” Consequently, to write  $\neg \exists x$  in front of this is to deny the existence of such a proof.

Now any such predicate expressible in our formal system has a Gödel number, and it is amusing to consider the cartoons in Figure 5.1.

In the first instance, we have the single-free-variable expression  $\neg \exists x \text{Proof}(x, y, y)$  preparing to eat  $y$ . Its own Gödel number is denoted by  $g$ . In the next instance, the character is given its own Gödel number to eat, and upon its ingestion the character is transformed to a predicate with no free variables — and hence no mouth. Naturally, even the resulting formula has a Gödel number,  $g'$ .

*Gödel's Theorem:*  $\neg \exists x \text{Proof}(x, g, g)$  is true but not provable in the formal arithmetic system.

The proof of this theorem takes only a few lines:

Suppose  $\neg \exists x \text{Proof}(x, g, g)$  is provable in the system, and let  $p$  be the Gödel number of its proof  $P$ . We then have

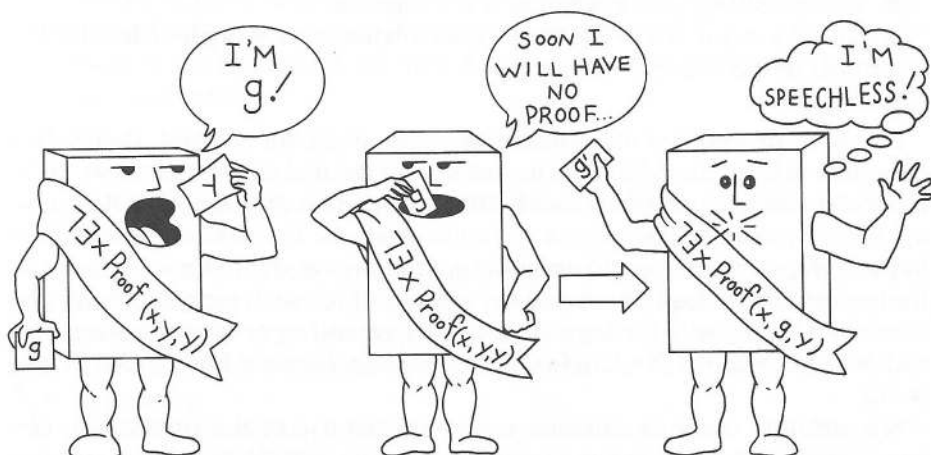


Figure 5.1 A certain formula ingests its own Gödel number

Proof( $p, g, g$ )

is true since  $P$  is a proof of  $G$  with  $g$  substituted for its one free variable. But, of course, Proof( $p, g, g$ ) contradicts  $\neg \exists x \text{Proof}(x, g, g)$ , and we are left with the conclusion that no such proof  $P$  exists.

The formula  $\neg \exists x \text{Proof}(x, g, g)$  is certainly true because we have just established the claim which it makes about itself—that it has no proof!

Such a proof, where a two-variable predicate is given the same value for both its arguments, is called a *proof by diagonalization*, and it crops up frequently in the theory of infinite sets and mathematical logic. Cantor was the first to use such an argument in showing that the real numbers are not countable.

Are there true statements which mathematicians are trying to prove at this very moment, but never will? What about Goldbach's conjecture which states that every even number is the sum of two primes? Certainly, no one has proved this statement so far, yet most mathematicians think it is true.

The struggle to formalize mathematics in a mechanical way led to the discovery of a basic and deeply seated problem in mathematics itself. The discovery was to be paralleled a few years later when the attempt to formalize "effective procedures" led to the discovery of a basic inadequacy in computers. There are some tasks just as impossible for computers (see Chapter 59) as for mathematicians.

## Problems

1. Write out the Gödel numbers for the integers 0, 1, 2, and 3.
2. Is it possible for two different expressions to have the same Gödel number? If so, give an example. If not, explain the impossibility.
3. What is the difference between our proof of Gödel's theorem and a proof in the formal arithmetic system? Can our proof ever be expressed as a proof in this system?

## References

- S. C. Kleene. *Introduction to Metamathematics*. Van Nostrand, Princeton, N.J., 1950.
- Raymond M. Smullyan. *Gödel's Incompleteness Theorem*. Oxford University Press, New York, 1992.