ABSTRACTIONS BLOG

How Gödel's Proof Works

By NATALIE WOLCHOVER

July 14, 2020

His incompleteness theorems destroyed the search for a mathematical theory of everything. Nearly a century later, we're still coming to grips with the consequences.

81 |

Every mathematical system will have some statements that can never be proved.

—

Olena Shmahalo/Quanta Magazine

In 1931, the Austrian logician Kurt Gödel pulled off arguably one of the most stunning intellectual achievements in history.

Mathematicians of the era sought a solid foundation for mathematics: a set of basic mathematical facts, or axioms, that was both consistent — never leading to contradictions — and complete, serving as the building blocks of all mathematical truths.

But Gödel's shocking incompleteness theorems, published when he was just 25, crushed that dream. He proved that any set of axioms you could posit as a possible foundation for math will inevitably be incomplete; there will always be true facts about numbers that cannot be proved by those axioms. He also showed that no candidate set of axioms can ever prove its own consistency.

His incompleteness theorems meant there can be no mathematical theory of everything, no unification of what's provable and what's true. What mathematicians can prove depends on their starting assumptions, not on any fundamental ground truth from which all answers spring.

In the 89 years since Gödel's discovery, mathematicians have stumbled upon just the kinds of unanswerable questions his theorems foretold. For example, Gödel himself helped establish that the continuum hypothesis, which concerns the sizes of infinity, is undecidable, as is the halting problem, which asks whether a computer program fed with a random input will run forever or eventually halt. Undecidable questions have even arisen in physics, suggesting that Gödelian incompleteness afflicts not just math, but — in some ill-understood way — reality.

Here's a simplified, informal rundown of how Gödel proved his theorems.

Gödel Numbering

Gödel's main maneuver was to map statements *about* a system of axioms onto statements *within* the system — that is, onto statements about numbers. This mapping allows a system of axioms to talk cogently about itself.

The first step in this process is to map any possible mathematical statement, or series of statements, to a unique number called a Gödel number.

The slightly modified version of Gödel's scheme presented by Ernest Nagel and James Newman in their 1958 book, *Gödel's Proof*, begins with 12 elementary symbols that serve as the vocabulary for expressing a set of basic axioms. For example, the statement that something exists can be expressed by the symbol ∃, while addition is expressed by +. Importantly, the symbol *s*, denoting "successor of," gives a way of specifying numbers; *ss*0, for example, refers to 2.

These twelve symbols then get assigned the Gödel numbers 1 through 12.

Constant sign	Gödel number	Usual Meaning
~	1	not
∨	2	or
⊃	3	if...then...
∃	4	there is an...
=	5	equals
0	6	zero
s	7	the successor of
(	8	punctuation mark
)	9	punctuation mark
,	10	punctuation mark
+	11	plus
×	12	times

Next, letters representing variables, starting with *x*, *y* and *z*, map onto prime numbers greater than 12 (that is, 13, 17, 19, ...).

Then any combination of these symbols and variables — that is, any arithmetical formula or sequence of formulas that can be constructed — gets its own Gödel number.

For example, consider 0 = 0. The formula's three symbols correspond to Gödel numbers 6, 5 and 6. Gödel needs to change this three-number sequence into a single, unique number — a number that no other sequence of symbols will generate. To do this, he takes the first three primes (2, 3 and 5), raises each to the Gödel number of the symbol in the same position in the sequence, and multiplies them together. Thus 0 = 0 becomes $2^6 \times 3^5 \times 5^6$, or 243,000,000.

The mapping works because no two formulas will ever end up with the same Gödel number. Gödel numbers are integers, and integers only factor into primes in a single way. So the only prime factorization of 243,000,000 is $2^6 \times 3^5 \times 5^6$, meaning there's only one possible way to decode the Gödel number: the formula 0 = 0.

Gödel then went one step further. A mathematical proof consists of a sequence of formulas. So Gödel gave every sequence of formulas a unique Gödel number too. In this case, he starts with the list of prime numbers as before — 2, 3, 5 and so on. He then raises each prime to the Gödel number of the formula at the same position in the sequence ($2^{243,000,000} \times \ldots$, if 0 = 0 comes first, for example) and multiplies everything together.

Arithmetizing Metamathematics

The real boon is that even statements *about* arithmetic formulas, called metamathematical statements, can themselves be translated into formulas with Gödel numbers of their own.

First consider the formula ~(0 = 0), meaning "zero does not equal zero." This formula is clearly false. Nevertheless, it has a Gödel number: 2 raised to the power of 1 (the Gödel number of the symbol ~), multiplied by 3 raised to the power of 8 (the Gödel number of the "open parenthesis" symbol), and so on, yielding $2^1 \times 3^8 \times 5^6 \times 7^5 \times 11^6 \times 13^9$.

Because we can generate Gödel numbers for all formulas, even false ones, we can talk sensibly about these formulas by talking about their Gödel numbers.

Consider the statement, "The first symbol of the formula ~(0 = 0) is a tilde." This (true) metamathematical statement about ~(0 = 0) translates into a statement about the formula's Gödel number — namely, that its first exponent is 1, the Gödel number for a tilde. In other words, our statement says that $2^1 \times 3^8 \times 5^6 \times 7^5 \times 11^6 \times 13^9$ has only a single factor of 2. Had ~(0 = 0) begun with any symbol other than a tilde, its Gödel number would have at least two factors of 2. So, more precisely, 2 is a factor of $2^1 \times 3^8 \times 5^6 \times 7^5 \times 11^6 \times 13^9$, but 2^2 is not a factor.

We can convert the last sentence into a precise arithmetical formula that we can write down* using elementary symbols. This formula of course has a Gödel number, which we could calculate by mapping its symbols onto powers of primes.

This example, Nagel and Newman wrote, "exemplifies a very general and deep insight that lies at the heart of Gödel's discovery: typographical properties of long chains of symbols can be talked about in an indirect but perfectly accurate manner by instead talking about the properties of prime factorizations of large integers."

Conversion into symbols is also possible for the metamathematical statement, "There exists some sequence of formulas with Gödel number x that proves the formula with Gödel number k" — or, in short, "The formula with Gödel number k can be proved." The ability to "arithmetize" this kind of statement set the stage for the coup.

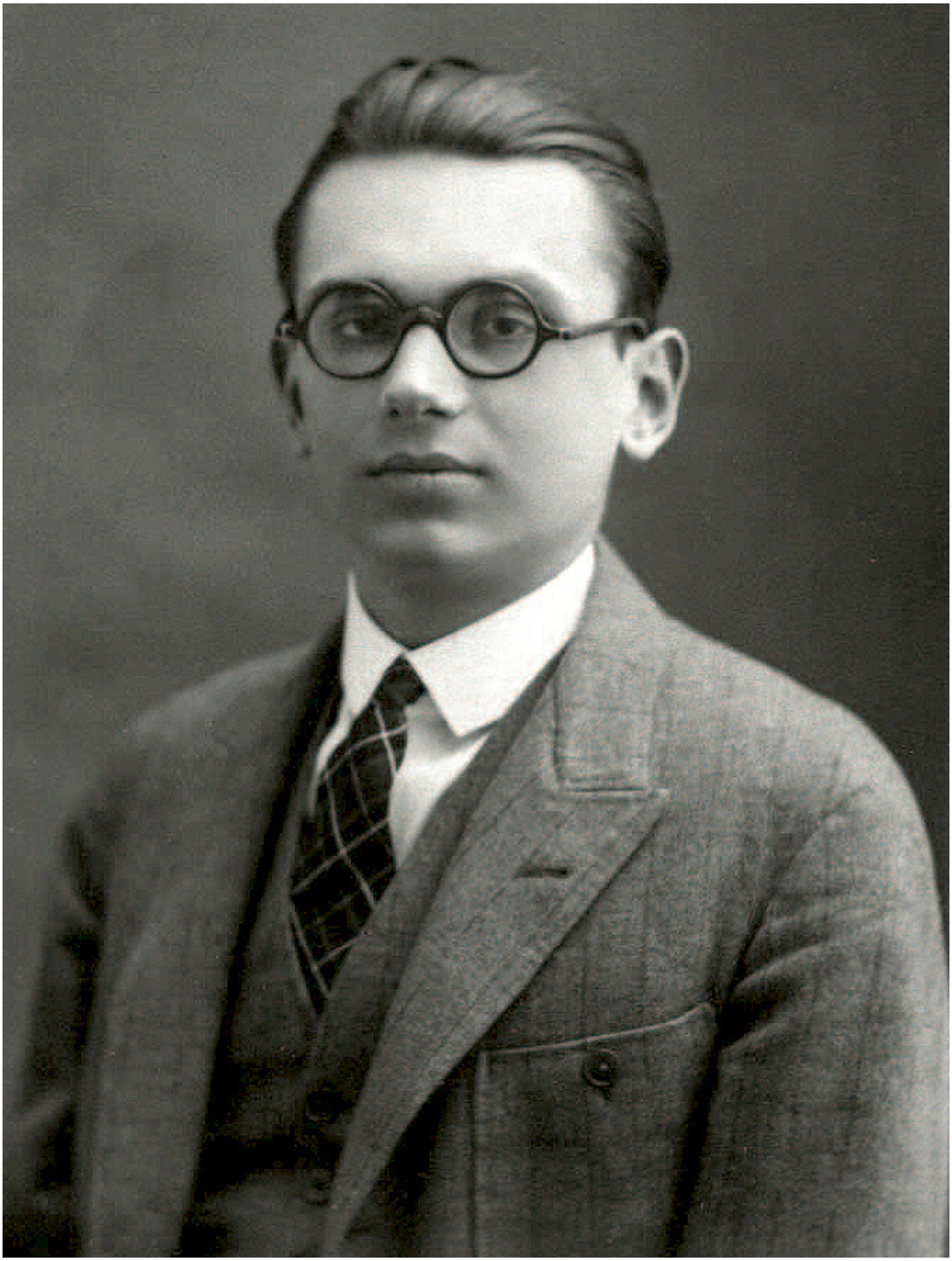
G Itself

Gödel's extra insight was that he could substitute a formula's own Gödel number in the formula itself, leading to no end of trouble.

To see how substitution works, consider the formula $(\exists x)(x = sy)$. (It reads, "There exists some variable x that is the successor of y," or, in short, "y has a successor.") Like all formulas, it has a Gödel number — some large integer we'll just call m.

Now let's introduce m into the formula in place of the symbol y. This forms a new formula, $(\exists x)(x = sm)$, meaning, "m has a successor." What shall we call this formula's Gödel number? There are three pieces of information to convey: We started with the formula that has Gödel number m. In it, we substituted m for the symbol y. And according to the mapping scheme introduced earlier, the symbol y has the Gödel number 17. So let's designate the new formula's Gödel number sub(m, m, 17).

Substitution forms the crux of Gödel's proof.



Kurt Gödel as a student in Vienna. He published his incompleteness theorems in 1931, a year after he graduated.

—

Kurt Gödel Papers, the Shelby White and Leon Levy Achives Center, Institute for Advanced Study

He considered a metamathematical statement along the lines of "The formula with Gödel number sub(y, y, 17) cannot be proved." Recalling the notation we just learned, the formula with Gödel number sub(y, y, 17) is the one obtained by taking the formula with Gödel number y (some unknown variable) and substituting this variable y anywhere there's a symbol whose Gödel number is 17 (that is, anywhere there's a y).

Things are getting trippy, but nevertheless, our metamathematical statement — "The formula with Gödel number sub(y, y, 17) cannot be proved" — is sure to translate into a formula with a unique Gödel number. Let's call it n.

Now, one last round of substitution: Gödel creates a new formula by substituting the number n anywhere there's a y in the previous formula. His new formula reads, "The formula with Gödel number sub(n, n, 17) cannot be proved." Let's call this new formula G.

Naturally, G has a Gödel number. What's its value? Lo and behold, it must be sub(n, n, 17). By definition, sub(n, n, 17) is the Gödel number of the formula that results from taking the formula with Gödel number n and substituting n anywhere there's a symbol with Gödel number 17. And G is exactly this formula! Because of the uniqueness of prime factorization, we now see that the formula G is talking about is none other than G itself.

G asserts of itself that it can't be proved.

But can G be proved? If so, this would mean there's some sequence of formulas that proves the formula with Gödel number sub(n, n, 17). But that's the opposite of G, which says no such proof exists. Opposite statements, G and ~G, can't both be true in a consistent axiomatic system. So the truth of G must be undecidable.

However, although G is undecidable, it's clearly true. G says, "The formula with Gödel number sub(n, n, 17) cannot be proved," and that's exactly what we've found to be the case! Since G is true yet undecidable within the axiomatic system used to construct it, that system is incomplete.

You might think you could just posit some extra axiom, use it to prove G, and resolve the paradox. But you can't. Gödel showed that the augmented axiomatic system will allow the construction of a new, true formula G′ (according to a similar blueprint as before) that can't be proved within the new, augmented system. In striving for a complete mathematical system, you can never catch your own tail.

No Proof of Consistency

We've learned that if a set of axioms is consistent, then it is incomplete. That's Gödel's first incompleteness theorem. The second — that no set of axioms can prove its own consistency — easily follows.

What would it mean if a set of axioms could prove it will never yield a contradiction? It would mean that there exists a sequence of formulas built from these axioms that proves the formula that means, metamathematically, "This set of axioms is consistent." By the first theorem, this set of axioms would then necessarily be incomplete.

But "The set of axioms is incomplete" is the same as saying, "There is a true formula that cannot be proved." This statement is equivalent to our formula G. And we know the axioms can't prove G.

So Gödel has created a proof by contradiction: If a set of axioms could prove its own consistency, then we would be able to prove G. But we can't. Therefore, no set of axioms can prove its own consistency.

Gödel's proof killed the search for a consistent, complete mathematical system. The meaning of incompleteness "has not been fully fathomed," Nagel and Newman wrote in 1958. It remains true today.

---

*For the curious, the statement reads: "There exists some integer x such that x multiplied by 2 is equal to $2^1 \times 3^8 \times 5^6 \times 7^5 \times 11^6 \times 13^9$, and there does not exist any integer x such that x multiplied by 4 is equal to $2^1 \times 3^8 \times 5^6 \times 7^5 \times 11^6 \times 13^9$." The corresponding formula is:

$$(\exists x)(x \times ss0 = sss \ldots sss0) \cdot \sim(\exists x)(x \times ssss0 = sss \ldots sss0)$$

where *sss … sss*0 stands for $2^1 \times 3^8 \times 5^6 \times 7^5 \times 11^6 \times 13^9$ copies of the successor symbol s. The symbol · means "and," and is shorthand for a longer expression in the fundamental vocabulary: $p \cdot q$ stands for $\sim(\sim p \vee \sim q)$. [Back to article.]

This article was reprinted on Wired.com.