inp.1 Introduction

inc:inp:int:sec

Hilbert thought that a system of axioms for a mathematical structure, such as the natural numbers, is inadequate unless it allows one to derive all true statements about the structure. Combined with his later interest in formal systems of deduction, this suggests that he thought that we should guarantee that, say, the formal systems we are using to reason about the natural numbers is not only consistent, but also *complete*, i.e., every statement in its language is either provable or its negation is. Gödel's first incompleteness theorem shows that no such system of axioms exists: there is no complete, consistent, axiomatizable formal system for arithmetic. In fact, no "sufficiently strong," consistent, axiomatizable mathematical theory is complete.

A more important goal of Hilbert's, the centerpiece of his program for the justification of modern ("classical") mathematics, was to find finitary consistency proofs for formal systems representing classical reasoning. With regard to Hilbert's program, then, Gödel's second incompleteness theorem was a much bigger blow. The second incompleteness theorem can be stated in vague terms, like the first incompleteness theorem. Roughly speaking, it says that no sufficiently strong theory of arithmetic can prove its own consistency. We will have to take "sufficiently strong" to include a little bit more than $\mathbf{Q}$.

The idea behind Gödel's original proof of the incompleteness theorem can be found in the Epimenides paradox. Epimenides, a Cretan, asserted that all Cretans are liars; a more direct form of the paradox is the assertion "this sentence is false." Essentially, by replacing truth with provability, Gödel was able to formalize a sentence which, in a roundabout way, asserts that it itself is not provable. If that sentence were provable, the theory would then be inconsistent. Assuming ω-consistency—a property stronger than consistency—Gödel was able to show that this sentence is also not refutable from the system of axioms he was considering.

The first challenge is to understand how one can construct a sentence that refers to itself. For every formula φ in the language of $\mathbf{Q}$, let $\ulcorner\varphi\urcorner$ denote the numeral corresponding to $^{\#}\varphi^{\#}$. Think about what this means: φ is a formula in the language of $\mathbf{Q}$, $^{\#}\varphi^{\#}$ is a natural number, and $\ulcorner\varphi\urcorner$ is a *term* in the language of $\mathbf{Q}$. So every formula φ in the language of $\mathbf{Q}$ has a *name*, $\ulcorner\varphi\urcorner$, which is a term in the language of $\mathbf{Q}$; this provides us with a conceptual framework in which formulas in the language of $\mathbf{Q}$ can "say" things about other formulas. The following lemma is known as the fixed-point lemma.

Lemma inp.1. *Let $\mathbf{T}$ be any theory extending $\mathbf{Q}$, and let $\psi(x)$ be any formula with only the variable x free. Then there is a sentence φ such that $\mathbf{T}$ proves $\varphi \leftrightarrow \psi(\ulcorner\varphi\urcorner)$.*

The lemma asserts that given any property $\psi(x)$, there is a sentence φ that asserts "$\psi(x)$ is true of me."

How can we construct such a sentence? Consider the following version of the Epimenides paradox, due to Quine:

> "Yields falsehood when preceded by its quotation" yields falsehood when preceded by its quotation.

This sentence is not directly self-referential. It simply makes an assertion about the syntactic objects between quotes, and, in doing so, it is on par with sentences like

1. "Robert" is a nice name.
2. "I ran." is a short sentence.
3. "Has three words" has three words.

But what happens when one takes the phrase "yields falsehood when preceded by its quotation," and precedes it with a quoted version of itself? Then one has the original sentence! In short, the sentence asserts that it is false.

Photo Credits

Bibliography