Gödel's Incompleteness Theorems

Aditya Dwarkesh [1] Satbhav Voleti [2]

[1]*Indian Institute of Science Education and Research, Kolkata* [2]*Indian Institute of Science Education and Research, Kolkata*

Introduction

Is it possible to formalize all of arithmetic in such a way that every mathematical statement is either provable or disprovable in a finite, mechanical sequence of steps?

Such a formalization was the ambitious goal of Whitehead & Russell's *Principia Mathematica*. Unfortunately for them, in 1931, Kurt Gödel proved and published a theorem which stated that such a system cannot exist.

Gödel went on to explicitly give an example of a mathematical statement which is neither provable nor disprovable in the relevant system. His two theorems are often considered to be the deepest results yet in the foundations of mathematics.

Indeed, even if one were to try to 'force' the statement in by adding it as an axiom of the system, it was shown that one could always construct a new "Gödel statement" for the fresh system. The difficulty was a fundamental one.

Background

First-order language: A first-order language is a set of symbols– variables, constants, predicates, functions, logical connectives and the universal quantifier. Recursive definitions are easy to give for 'meaningful' strings of these symbols, namely well-formed formulae. A well-formed formulae with no free variable is called a sentence.

First-order formal system: A first-order formal system includes a set of axioms (all well-formed formulae of a certain form) and a set of rules of deduction. A finite sequence of well-formed formulae starting from ϕ_1 and ending at ϕ_2 is a derivation of the latter from the former if each of these steps is either an axiom or follows from a rule of deduction.

Theory: A theory is a deductively closed set of sentences.

Completeness: A first-order theory is complete if for every sentence σ, either σ or $\neg\sigma$ is provable in it.

Consistency: A first-order theory is consistent if for no formula ϕ, both ϕ and $\neg\phi$ are provable in it.

For the purpose of this theorem, the first order theory of interest is Peano Arithmetic.

ω-consistency: A set of sentences is said to be ω-consistent if for all formulae of type $\phi(x)$ with one free variable, it is not the case that both $\neg\forall\neg\phi(x)$ and $\neg\phi(n)$ for all naturals n.

Recursive functions: A function from $\mathbb{N}^n$ to $\mathbb{N}$ is said to be *recursive* if it is a constant, the successor, the projection, a composition or minimization or primitive recursion of recursive functions.

Representability: A function $f : \mathbb{N}^n \to \mathbb{N}$ is said to be representable in a theory T iff there exists a formula $A(x_1, \ldots, x_{n+1})$ such that $T \vdash \forall x A(a_1, \ldots, a_n, x) \leftrightarrow (x = b)$ whenever $f(a_1, \ldots, a_n) = b$.

The first theorem

1. Every recursive relation is representable in $\mathbf{T}$.
2. Every formula in $\mathbf{T}$ can be arithmetized, such that for each formula corresponds a unique integer. This can be done by encoding each symbol as a unique integer and then a sequence of symbols via $p_1^{a_1} * \ldots p_n^{a_n}$, where p_i is the i^{th} prime and a_i is the code of the i^{th} symbol. We call the integer corresponding to a formula its *Gödel number*, and denote it (for a formula ϕ) by $\ulcorner\phi\urcorner$.
3. $\mathscr{P}_{\mathbf{T}}$ defined in the following manner is recursive (and thus representable in T): (x, y) $\in \mathscr{P}_{\mathbf{T}}$ if and only if the formula associated with the integer y constitutes a proof for the formula associated with the integer x. The representation of this relation in T is called the "provability predicate", and shall be denoted by $\mathscr{P}_{\mathbf{T}}$.
4. Let $F(x)$ be a formula with one free variable. There exists a sentence ϕ such that $\phi \leftrightarrow F(\ulcorner\phi\urcorner)$.
5. Let $\mathscr{B}_{\mathbf{T}}(\mathbf{x})$ be the formula $\exists p \mathscr{P}_{\mathbf{T}}(\mathbf{x}, \mathbf{p})$. Informally, $\mathscr{B}_T(x)$ reads 'There exists a proof of x.' By the above, there exists a statement $\mathscr{G}$ such that $\mathbf{T} \vdash \mathscr{G} \leftrightarrow \neg\mathscr{B}_{\mathbf{T}}(\ulcorner\mathscr{G}\urcorner)$. We claim that neither $\mathscr{G}$ nor $\neg\mathscr{G}$ are provable in $\mathbf{T}$.
6. Suppose $\mathscr{G}$. $\mathscr{G} \implies \mathscr{B}_{\mathbf{T}}(\ulcorner\mathscr{G}\urcorner)$, by the Hilbert-Bernays theorems (informally, this says that if G is provable, then 'G is provable' is provable). But also, $\neg\mathscr{B}_{\mathbf{T}}(\ulcorner\mathscr{G}\urcorner)$, contradicting the ω-consistency of $\mathbf{T}$.
7. Suppose $\neg\mathscr{G}$. We will use another Hilbert-Bernays theorem, and call it HB (2): $\mathscr{B}_{\mathbf{T}}(\ulcorner A \to B\urcorner) \to (\mathscr{B}_{\mathbf{T}}(\ulcorner A\urcorner) \to \mathscr{B}_{\mathbf{T}}(\ulcorner B\urcorner))$. Informally, this says that if $A \implies B$ is provable, then 'A is provable' implies 'B is provable'.
   Finally, let $\perp$ stand for your favourite contradiction. We will show that our assumption entails the provability of a contradiction in $\mathbf{T}$, contradicting its ω-consistency.

1.	$\mathbf{T} \vdash \mathscr{B}_{\mathbf{T}}(\ulcorner\mathscr{G}\urcorner)$	by $\mathscr{G}$ being a fixed point
2.	$\mathbf{T} \vdash \mathscr{B}_{\mathbf{T}}(\ulcorner\neg\mathscr{G}\urcorner)$	Hilbert-Bernays
3.	$\mathbf{T} \vdash \neg\mathscr{G} \to (\mathscr{G} \to\perp)$	Tautology
4.	$\mathbf{T} \vdash \mathscr{B}_{\mathbf{T}}(\ulcorner\neg\mathscr{G} \to (\mathscr{G} \to\perp)\urcorner)$	Hilbert-Bernays
5.	$\mathbf{T} \vdash \mathscr{B}_{\mathbf{T}}(\ulcorner\neg\mathscr{G}\urcorner) \to \mathscr{B}_{\mathbf{T}}(\ulcorner\mathscr{G} \to\perp\urcorner))$	HB (2) and Modus Ponens
6.	$\mathbf{T} \vdash \mathscr{B}_{\mathbf{T}}(\ulcorner\mathscr{G} \to\perp\urcorner)$	Modus Ponens
7.	$\mathbf{T} \vdash \mathscr{B}_{\mathbf{T}}(\ulcorner\mathscr{G}\urcorner) \to \mathscr{B}_{\mathbf{T}}(\ulcorner\perp\urcorner)$	HB (2)
8.	$\mathbf{T} \vdash \mathscr{B}_{\mathbf{T}}(\ulcorner\perp\urcorner)$	Modus Ponens

Informally, $\mathscr{G}$ may be said to express the Liar's Paradox: It reads, 'This sentence is true but unprovable.'

A word of caution

Gödel's theorems do not, in fact, hold for all formal systems. Tarski's axiomatization of elementary Euclidean geometry is an example of a formal system which is both consistent and complete!

The catch is that our theory needs to be sufficiently complex, in some appropriate sense, for the theorems to apply. In particular, every recursive relation must be representable in it. This is not true of Tarski's system; on the other hand, a formal first-order system capable of doing reasonably interesting arithmetic will satisfy this.

Nor is it true of second-order or other such variants where quantification over predicates is allowed. Using second-order axiomatization, a complete, consistent and decidable theory of the real closed fields was given by Tarski.

The second theorem

Gödel's second theorem states that no consistent extension of arithmetic can prove its own consistency. Its proof is based on Löb's theorem, which states that if $\mathscr{B}_{\mathbf{T}}(\ulcorner\sigma\urcorner) \to \sigma$, then σ (for any sentence σ).
Assuming this, let $\mathbf{T}$ be any recursively axiomatized extension of Peano arithmetic. Note that $\mathscr{B}_{\mathbf{T}}(\ulcorner\perp\urcorner) \to \perp$ is equivalent to $\neg\mathscr{B}_{\mathbf{T}}(\ulcorner\perp\urcorner)$, a sentence which expresses the consistency of the theory. Now, by Löb's theorem, if $\mathscr{B}_{\mathbf{T}}(\ulcorner\perp\urcorner) \to \perp$, then $\perp$. Hence, we must have that if $\neg\mathscr{B}_{\mathbf{T}}(\ulcorner\perp\urcorner)$, then $\perp$. The contrapositive of this is the required statement.

Following is a proof of Löb's theorem:

1.	$\mathbf{T} \vdash B \leftrightarrow (\mathscr{B}_{\mathbf{T}}(\ulcorner B\urcorner) \to \sigma)$
2.	$\mathbf{T} \vdash \mathscr{B}_{\mathbf{T}}(\ulcorner B \to (\mathscr{B}_{\mathbf{T}}(\ulcorner B\urcorner) \to \sigma)\urcorner)$
3.	$\mathbf{T} \vdash \mathscr{B}_{\mathbf{T}}(\ulcorner B\urcorner) \to \mathscr{B}_{\mathbf{T}}(\ulcorner\mathscr{B}_{\mathbf{T}}(\ulcorner B\urcorner) \to \sigma)\urcorner)$
4.	$\mathbf{T} \vdash \mathscr{B}_{\mathbf{T}}(\ulcorner B\urcorner) \to (\mathscr{B}_{\mathbf{T}}(\ulcorner\mathscr{B}_{\mathbf{T}}(\ulcorner B\urcorner)\urcorner) \to \mathscr{B}_{\mathbf{T}}(\ulcorner\sigma\urcorner))$
5.	$\mathbf{T} \vdash \mathscr{B}_{\mathbf{T}}(\ulcorner B\urcorner) \to \mathscr{B}_{\mathbf{T}}(\ulcorner\mathscr{B}_{\mathbf{T}}(\ulcorner B\urcorner)\urcorner)$
6.	$\mathbf{T} \vdash (\mathscr{B}_{\mathbf{T}}(\ulcorner B\urcorner) \to \mathscr{B}_{\mathbf{T}}(\ulcorner\mathscr{B}_{\mathbf{T}}(\ulcorner B\urcorner)\urcorner)) \to (\mathscr{B}_{\mathbf{T}}(\ulcorner B\urcorner) \to \mathscr{B}_{\mathbf{T}}(\ulcorner\sigma\urcorner))$
7.	$\mathbf{T} \vdash \mathscr{B}_{\mathbf{T}}(\ulcorner B\urcorner) \to \mathscr{B}_{\mathbf{T}}(\ulcorner\sigma\urcorner)$
8.	$\mathbf{T} \vdash \mathscr{B}_{\mathbf{T}}(\ulcorner\sigma\urcorner) \to \sigma$
9.	$\mathbf{T} \vdash (\mathscr{B}_{\mathbf{T}}(\ulcorner\sigma\urcorner) \to \sigma) \to (\mathscr{B}_{\mathbf{T}}(\ulcorner B\urcorner) \to (\mathscr{B}_{\mathbf{T}}(\ulcorner\sigma\urcorner) \to \sigma))$
10.	$\mathbf{T} \vdash \mathscr{B}_{\mathbf{T}}(\ulcorner B\urcorner) \to (\mathscr{B}_{\mathbf{T}}(\ulcorner\sigma\urcorner) \to \sigma)$
11.	$\mathbf{T} \vdash (\mathscr{B}_{\mathbf{T}}(\ulcorner B\urcorner) \to \mathscr{B}_{\mathbf{T}}(\ulcorner\sigma\urcorner)) \to (\mathscr{B}_{\mathbf{T}}(\ulcorner B\urcorner) \to \sigma)$
12.	$\mathbf{T} \vdash \mathscr{B}_{\mathbf{T}}(\ulcorner B\urcorner) \to \sigma$
13.	$\mathbf{T} \vdash B$
14.	$\mathbf{T} \vdash \mathscr{B}_{\mathbf{T}}(\ulcorner B\urcorner)$
15.	$\mathbf{T} \vdash \sigma$

Aftermath

Hilbert's formalist program for mathematics was an ambitious project to axiomatize mathematics such that it retained three main characteristics: (1) finiteness–all results about infinitary objects (like irrationals) must be proven using formalism of finitary objects, (2) completeness–all true mathematical statements are provable, (3) consistency–no contradiction is obtainable in the formalism. Rather than think of strange mathematical objects as 'real', he decided to only commit to 'finitary' objects and the rest are to be thought of as meaningless symbols to be jumbled around according to our choice of rules or formalism.

Gödel's theorems showed that completeness is not possible with just finitary means and that consistency cannot be proven by the formalism. It delivered a fatal blow to Hilbert's optimistic vision.

References

[1] Kurt Gödel, Bernard Meltzer, and Richard Schlegel. On formally undecidable propositions of principia mathematica and related systems. 1962.

[2] Ernest Nagel and James R. Newman. *Godel's Proof*. New York, NY, USA: Routledge, 1958.