

Proof of the Incompleteness Theorem

We bring together all the tools developed so far to prove Gödel's First Incompleteness Theorem.

Note: This is a good time to reread the chapter “An Overview of Gödel's Incompleteness Theorem” and also the preceding chapter, “Diagonalization.”

1. The incompleteness of Peano arithmetic

1.1. Statement

Theorem 1.

If $\mathcal{PA}$ is consistent, then $\mathcal{PA}$ is incomplete; that is, there exists a closed formula $\mathcal{G}$ such that:

$$\mathcal{PA} \not\vdash \mathcal{G} \quad \text{and} \quad \mathcal{PA} \not\vdash \neg\mathcal{G}$$

Thus there exist statements that are true (in the standard model $\mathbb{N}$) but unprovable.

- Recall that $\mathcal{PA}$ denotes the theory of Peano arithmetic, that is, the usual axioms associated with the natural numbers.
- Incompleteness means that there exists a closed formula $\mathcal{G}$ that is not provable from the Peano axioms and whose negation $\neg\mathcal{G}$ is not provable either.
- Since $\mathcal{G}$ is a closed formula, in every interpretation—for example, the model $\mathbb{N}$ or another model—exactly one of the following two possibilities holds: either $\mathcal{G}$ is true or $\neg\mathcal{G}$ is true. One of them is true, and both are unprovable.
- Thus, for the standard model $\mathbb{N}$, and for every other model, there exist statements that are true (in that model) and unprovable (in $\mathcal{PA}$).

1.2. Consistency and incompleteness

Review of consistency.

- A theory $\mathcal{T}$ is **inconsistent** if there exists a formula $\mathcal{Q}$ such that:

$$\mathcal{T} \vdash \mathcal{Q} \quad \text{and} \quad \mathcal{T} \vdash \neg\mathcal{Q}.$$

It is called **consistent** otherwise.

- Equivalently, $\mathcal{T}$ is inconsistent if $\mathcal{T} \vdash \perp$, in which case $\mathcal{T}$ proves every formula (and its negation) by the rule of explosion. An inconsistent theory is therefore of no interest.

- If $\mathcal{T}$ has a model (that is, $\models_S \mathcal{T}$ for some structure S), then $\mathcal{T}$ is consistent by the Soundness Theorem.

Review of completeness.

- A theory $\mathcal{T}$ is **incomplete** if there exists a closed formula Q such that:

$$\mathcal{T} \not\vdash Q \quad \text{and} \quad \mathcal{T} \not\vdash \neg Q.$$

- It is called **complete** otherwise; that is, for every closed formula Q , we have:

$$\mathcal{T} \vdash Q \quad \text{or} \quad \mathcal{T} \vdash \neg Q.$$

In the statement of Gödel's theorem, we start with a consistent theory and prove that it is incomplete.

1.3. The consistency assumption

When we stated the Incompleteness Theorem in the chapter “An Overview of Gödel's Incompleteness Theorem,” we did not state the assumption “ $\mathcal{PA}$ is consistent.” Why not? This is one of the subtleties in the statement.

If we work in a world in which the natural numbers $\mathbb{N}$ exist (for example, in $\mathcal{ZFC}$), then $\mathbb{N}$ is a model of $\mathcal{PA}$ (everything is designed that way: the axioms of $\mathcal{PA}$ correspond to properties of the natural numbers), so $\mathcal{PA}$ is consistent, making the assumption unnecessary.

But if we wish to remain strictly within the framework of the language $\mathcal{L}_0$ and the theory $\mathcal{PA}$, then the existence of $\mathbb{N}$ is not given, and consequently neither is the consistency of $\mathcal{PA}$.

This issue is in fact unavoidable: the Second Incompleteness Theorem will state that a (sufficiently strong) theory cannot prove its own consistency.

2. ω -consistency, being a proof, and being provable

2.1. ω -consistency

We will prove Gödel's theorem under a slightly stronger assumption: ω -consistency (the assumption used in Gödel's original article). A little later, we will discuss how to dispense with it.

Definition.

A theory $\mathcal{T}$ is **ω -inconsistent** if there exists a formula $Q(x)$ (with x as its only free variable) such that:

$$\mathcal{T} \vdash \exists x Q(x)$$

$$\text{and for every } n \in \mathbb{N}, \quad \mathcal{T} \vdash \neg Q(\bar{n})$$

Otherwise it is called **ω -consistent**.

- This is therefore a type of inconsistency specific to the natural numbers: we have $\neg Q(0)$, $\neg Q(1)$, $\neg Q(2)$, and so on, and yet there exists an x such that $Q(x)$.
- In an ω -consistent theory, no such formula exists.
- This definition goes beyond the language $\mathcal{L}_0$ because it involves the natural numbers $n = 0$, $n = 1$, $\dots$
- If $\mathcal{T}$ has $\mathbb{N}$ as a model, then $\mathcal{T}$ is ω -consistent. Hence, if we regard $\mathbb{N}$ as a model of $\mathcal{PA}$, then $\mathcal{PA}$ is ω -consistent.

Here is the connection with ordinary consistency:

- If $\mathcal{T}$ is ω -consistent, then $\mathcal{T}$ is consistent.
- If $\mathcal{T}$ is inconsistent, then $\mathcal{T}$ is ω -inconsistent.

Justification: if $\mathcal{T}$ is inconsistent, then $\mathcal{T}$ proves everything, and therefore proves the formulas in the definition of ω -inconsistency. The other direction follows by contraposition.

These two notions are not equivalent, but a little later we will see Rosser's trick, which makes it possible to replace Gödel's assumption of " ω -consistency" with the weaker assumption of "consistency."

2.2. Being a proof and being provable

We return to the fundamental distinction between checking that something is a proof and deciding whether a formula is provable.

- Recall that $\text{is_proof}(N, n)$, which decides whether the sequence of formulas with super-Gödel number N constitutes a proof of the formula whose Gödel number is n , is a primitive recursive predicate.
- Therefore, by representability, there is a corresponding formula in $\mathcal{L}_0$:

$$\mathcal{I}s\text{-proof}(y, x)$$

- On the other hand, $\text{is_provable}(n)$ is not a primitive recursive predicate, because the search for a proof must range over an unbounded set of integers N .
- But this is not a problem, because in $\mathcal{L}_0$ it is perfectly legitimate to define the formula:

$$\mathcal{I}s\text{-provable}(x) : \exists y \mathcal{I}s\text{-proof}(y, x)$$

Note that to obtain this formula of $\mathcal{L}_0$, we used the representability of the relation "being a proof" as an intermediate step.

Corollary 1.

The set of provable formulas is definable.

Indeed, by definition, a formula $\mathcal{P}$ is provable (in $\mathcal{PA}$) if and only if $\mathcal{I}s\text{-provable}(\ulcorner \mathcal{P} \urcorner)$ is true (in $\mathbb{N}$).

This stands in contrast to the property "being true," which is not definable (see Tarski's theorem in the chapter "Diagonalization").

Remember these two crucial differences:

- "Being a proof" and "being provable" are fundamentally different predicates: one is easy to compute, while the other is not.
- "Being provable" and "being true" are also fundamentally different concepts: one is definable, while the other is not.

3. Proof of the Incompleteness Theorem for $\mathcal{PA}$

Here is the statement that we will prove:

Theorem 2.

If $\mathcal{PA}$ is ω -consistent, then $\mathcal{PA}$ is incomplete.

3.1. The Gödel formula

Recall that the formula $\text{Is-proof}(y, x)$ represents in $\mathcal{PA}$ the predicate “ y is the code of a proof of the formula with code x .” We then define:

$$\text{Is-provable}(x) : \exists y \text{ Is-proof}(y, x)$$

Consider the formula

$$\mathcal{P}(x) : \neg \text{Is-provable}(x)$$

By the diagonal lemma, there exists a closed formula $\mathcal{G}$, called the *Gödel formula*, such that

$$\mathcal{PA} \vdash (\mathcal{G} \leftrightarrow \mathcal{P}(\ulcorner \mathcal{G} \urcorner))$$

(We use $\ulcorner \mathcal{G} \urcorner$ both for the Gödel number of $\mathcal{G}$ and for its numeral.)

Thus, in $\mathcal{PA}$, the formula $\mathcal{G}$ states that it is not provable. In other words, it says: “I am not provable.” We will now show that this formula $\mathcal{G}$ is not provable and that $\neg \mathcal{G}$ is not provable either.

3.2. Proof that $\mathcal{PA} \not\vdash \mathcal{G}$

Suppose for a contradiction that $\mathcal{PA} \vdash \mathcal{G}$. Then there exists a proof of $\mathcal{G}$ from the axioms of $\mathcal{PA}$. Let N denote the super-Gödel number associated with the formulas that constitute this proof.

Thus $\text{is_proof}(N, \ulcorner \mathcal{G} \urcorner)$ holds (as a primitive recursive predicate), and therefore:

$$\begin{aligned} \mathcal{PA} &\vdash \text{Is-proof}(\overline{N}, \ulcorner \mathcal{G} \urcorner) \\ \text{therefore } \mathcal{PA} &\vdash \exists y \text{ Is-proof}(y, \ulcorner \mathcal{G} \urcorner) \\ \text{therefore } \mathcal{PA} &\vdash \text{Is-provable}(\ulcorner \mathcal{G} \urcorner) \\ \text{therefore } \mathcal{PA} &\vdash \neg \mathcal{G} \text{ by the diagonal equivalence.} \end{aligned}$$

Conclusion: if $\mathcal{PA} \vdash \mathcal{G}$, then $\mathcal{PA} \vdash \neg \mathcal{G}$, which implies that $\mathcal{PA}$ is inconsistent. But we assumed that $\mathcal{PA}$ is ω -consistent and hence consistent (this weaker assumption is enough for this first step). We obtain a contradiction.

3.3. Proof that $\mathcal{PA} \not\vdash \neg \mathcal{G}$

Suppose for a contradiction that $\mathcal{PA} \vdash \neg \mathcal{G}$. Then, by the diagonal equivalence, we have:

$$\mathcal{PA} \vdash \exists y \text{ Is-proof}(y, \ulcorner \mathcal{G} \urcorner)$$

If we define $Q(y)$ by $\text{Is-proof}(y, \ulcorner \mathcal{G} \urcorner)$, this can be written:

$$\mathcal{PA} \vdash \exists y Q(y)$$

On the other hand, since we assume $\mathcal{PA} \vdash \neg \mathcal{G}$ and $\mathcal{PA}$ is consistent (because it is ω -consistent), we have $\mathcal{PA} \not\vdash \mathcal{G}$ (which we in fact already proved in the preceding section).

This means that for every integer $N = 0, 1, 2, \dots$, it is false that N is a proof of $\ulcorner \mathcal{G} \urcorner$; in other words, $\neg \text{is_proof}(N, \ulcorner \mathcal{G} \urcorner)$ is true. By the (strong) representability of primitive recursive predicates, we have:

$$\mathcal{PA} \vdash \neg \text{Is-proof}(\overline{N}, \ulcorner \mathcal{G} \urcorner)$$

This means that for $N = 0, 1, 2, \dots$, we have

$$\mathcal{PA} \vdash \neg Q(\bar{N})$$

This implies that $\mathcal{PA}$ is ω -inconsistent, contradicting our assumption.

3.4. Is $\mathcal{G}$ true?

We have proved $\mathcal{PA} \not\vdash \mathcal{G}$ and $\mathcal{PA} \not\vdash \neg\mathcal{G}$. Suppose we work in a model of $\mathcal{PA}$. Then:

- either the sentence $\mathcal{G}$ is true and is not provable,
- or the sentence $\neg\mathcal{G}$ is true and is not provable either.

In either case, there are unprovable truths!

Let us work in the usual setting in which $\mathbb{N}$ is regarded as a model of $\mathcal{PA}$ (for example, within $\mathcal{ZFC}$). On the one hand, $\mathcal{PA}$ is ω -consistent (and therefore consistent). On the other hand, $\mathcal{G}$ is true! Indeed, $\mathcal{G}$ is not provable, which means that the formula $\neg\text{Is-provable}(\ulcorner \mathcal{G} \urcorner)$ is true (in $\mathbb{N}$). Therefore, by the diagonal equivalence, $\mathcal{G}$ is true.

Once again, this assumes that $\mathbb{N}$ is a model of $\mathcal{PA}$ (an assumption that lies outside the framework of Peano arithmetic). This is why Gödel states his theorem with a consistency assumption and without a notion of truth, so as to remain strictly within the framework of $\mathcal{PA}$.

3.5. Rosser's trick

Rosser (1936) replaces Gödel's assumption of ω -consistency with the weaker assumption " $\mathcal{PA}$ is consistent." Instead of using the sentence "I am not provable" in the proof, he uses the sentence: "If I am provable, then there exists a shorter proof of my negation."

Formally, the formula $\mathcal{P}(x)$ to which he applies the diagonal lemma is:

$$\mathcal{P}(x) : \quad \forall y \left(\text{Is-proof}(y, x) \rightarrow \exists z \leq y \text{ Is-proof}(z, \text{neg}(x)) \right)$$

I.e., it says: "If I am provable, then there exists a proof of my negation whose code is no greater than the code of my proof."

If $x = \ulcorner \mathcal{Q} \urcorner$, we write $\text{neg}(x) = \ulcorner \neg\mathcal{Q} \urcorner$.

Notice the bounded quantifier " $\exists z \leq y$ ", which avoids the need to introduce ω -consistency and enumerate $N = 0, 1, 2, \dots$ all the way to infinity.

4. The general Incompleteness Theorem

4.1. Statement

Theorem 3.

Let $\mathcal{T}$ be a consistent theory with recursive axioms that contains $\mathcal{PA}$. Then $\mathcal{T}$ is incomplete.

In other words, in any sufficiently strong theory, there are true but unprovable formulas.

Since Peano arithmetic is incomplete, one might think that it would be enough to add an unprovable formula (such as $\mathcal{G}$) to its axioms in order to make it immediately provable in this new theory. But Gödel's theorem proves that this new theory will also be incomplete: there will be a new unprovable formula. There is therefore no way out: there will always be true but unprovable formulas.

Let us revisit the assumptions. We have already recalled what it means for " $\mathcal{T}$ to be consistent." The assumption " $\mathcal{T}$ contains $\mathcal{PA}$ " means exactly " $\mathcal{PA} \subset \mathcal{T}$ "; that is, the Peano axioms are among the axioms of $\mathcal{T}$. Informally, we say that $\mathcal{T}$ is "sufficiently strong." Being able to perform computations with integers is, however, a fairly weak requirement in mathematics.

It remains to explain the assumption “the axioms of $\mathcal{T}$ are recursive.”

4.2. Recursive functions: Church’s thesis

We will explain, without giving a complete definition, the notion of “recursive,” which is broader than “primitive recursive.”

Informal definition. A function $f : \mathbb{N}^p \rightarrow \mathbb{N}$ is a **recursive function** if it is computable.

The class of recursive functions is therefore larger than the class of primitive recursive functions, which were the “easily computable” functions.

There is, of course, a precise definition of a recursive function. It is a more technical variant of the notion of a primitive recursive function, with unbounded operations added. For example, the Ackermann function (see the chapter “Primitive Recursive Functions”) is recursive but not primitive recursive.

Church’s thesis is the heuristic claim that all computable functions are recursive (and conversely).

The result that matters to us is that every recursive function is representable (as with primitive recursive functions).

Let us finish our definitions:

- A set $A \subset \mathbb{N}$ is a **recursive set** (respectively, a primitive recursive set) if its characteristic function $\chi_A : \mathbb{N} \rightarrow \mathbb{N}$ is recursive (respectively, primitive recursive).
- A theory $\mathcal{T}$ has **recursive axioms** (respectively, primitive recursive axioms) if $A = \{\ulcorner \mathcal{P} \urcorner \mid \mathcal{P} \text{ is an axiom of } \mathcal{T}\}$, the set of Gödel numbers of the axioms of $\mathcal{T}$, is recursive (respectively, primitive recursive).

Requiring the set of axioms of a theory to be recursive simply means that we can determine whether or not a given formula is an axiom. This is an extremely reasonable requirement: we must be able to recognize the axioms of our theory!

4.3. Proof of the Incompleteness Theorem

Since we have not given the exact definition of a recursive function, we replace this assumption with the slightly stronger assumption “the axioms of $\mathcal{T}$ are primitive recursive,” which means that we can (easily) determine whether or not a given formula is an axiom.

This makes it possible to construct the primitive recursive predicate $\text{is_proof}(N, n) : \mathbb{N}^2 \rightarrow \{T, F\}$, and then represent it by a formula of $\mathcal{L}_0$, $\mathcal{I}s\text{-proof}(y, x)$. We also replace the consistency assumption with the stronger assumption of ω -consistency.

The proof is then exactly the same as for the theory $\mathcal{PA}$: we construct a formula $\mathcal{G}$ by diagonalization and prove that neither $\mathcal{G}$ nor $\neg\mathcal{G}$ is provable.

If $\mathbb{N}$ is a model of the theory $\mathcal{T}$, then $\mathcal{G}$ is true, although unprovable.

Returning to completeness.

By the Completeness Theorem, we know that if a formula is true in every model, then it is provable. This does not contradict the Incompleteness Theorem at all; it simply means here that there are models in which $\mathcal{G}$ is false. These models are called *nonstandard models* of $\mathcal{PA}$: they satisfy the Peano axioms but are not equal to $\mathbb{N}$ (for example, they contain infinitely large integers).