

An Overview of Gödel's Incompleteness Theorems

Before embarking on the construction of the tools needed to state Gödel's theorems precisely and prove them, we give an overview of the Incompleteness Theorems and the long road ahead.

1. Statements of the Theorems

1.1. Statement for the Arithmetic of Natural Numbers

Here is a still informal version of the First Incompleteness Theorem.

Theorem 1.

In the theory of the natural numbers, there are true but unprovable formulas.

We denote by $\mathcal{PA}$ the theory of arithmetic of the natural numbers. It is a list of formulas called the *Peano axioms* (for example, " $x + 0 = x$ " and " $x \times 0 = 0$ "). They are constructed so that the natural numbers $\mathbb{N}$ naturally satisfy these axioms.

In fact, Gödel did not speak of truth in his statement, in order to avoid discussions involving statements like the liar paradox, which asserts, "I am a liar." Instead, he constructs a formula $\mathcal{G}$ such that neither $\mathcal{G}$ nor $\neg\mathcal{G}$ is provable. The theory is then said to be *incomplete* (hence the name of the theorem).

Now, if we work in the model $\mathbb{N}$ of the natural numbers, which satisfies the Peano axioms, then either $\mathcal{G}$ is true or $\neg\mathcal{G}$ is true, so one of the two formulas is indeed true but unprovable.

1.2. General Theorem

One might think that the Peano axioms are to blame because they are poorly chosen or too few in number. But Gödel's theorem is actually more general.

Theorem 2.

Every consistent theory $\mathcal{T}$ that contains $\mathcal{PA}$ and has recursive axioms is incomplete.

Let us forget the hypotheses for a moment and return to Peano arithmetic. Since this theory is incomplete, there is a formula $\mathcal{P}_0$ that is not provable in $\mathcal{PA}$. Let $\mathcal{T}_0 = \mathcal{PA}$, and let $\mathcal{T}_1 = \mathcal{T}_0 \cup \{\mathcal{P}_0\}$ be the new theory obtained by adding the unprovable formula $\mathcal{P}_0$ to $\mathcal{T}_0$.

Now $\mathcal{P}_0$ is provable in $\mathcal{T}_1$: $\mathcal{T}_1 \vdash \mathcal{P}_0$, because $\mathcal{P}_0$ is one of the formulas in $\mathcal{T}_1$. Unfortunately, the general version of the Incompleteness Theorem states that this new theory $\mathcal{T}_1$ is also incomplete: there is a formula $\mathcal{P}_1$ such that neither $\mathcal{P}_1$ nor $\neg\mathcal{P}_1$ is provable. We could add $\mathcal{P}_1$ (or $\neg\mathcal{P}_1$) to construct a theory $\mathcal{T}_2, \dots$ but this only postpones the problem indefinitely.

Thus, no matter how many unprovable formulas we add one at a time as axioms, there will always be unprovable formulas. That is incompleteness!

1.3. Concepts Involved

Let us review the definitions of the terms appearing in the hypotheses and the conclusion.

Completeness.

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

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

- If we work in a *model* of $\mathcal{T}$, that is, a structure in which all the formulas of $\mathcal{T}$ are satisfied, then one of the statements Q or $\neg Q$ is true but not provable.

Consistency.

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

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

It is said to be *consistent* otherwise.

- Note carefully that the definition of “inconsistent” is not the negation of “incomplete”, because the negation of an “and” is an “or.”
- Thus, an inconsistent theory proves $\mathcal{T} \vdash (Q \wedge \neg Q)$, so $\mathcal{T} \vdash \perp$. Then, by the rule of explosion, $\mathcal{T} \vdash Q$ (whatever the formula Q may be), so we also have $\mathcal{T} \vdash \neg Q$.
- An inconsistent theory is therefore of no interest, since it cannot have a model. Indeed, if it had a model, then because $\mathcal{T} \vdash Q$ and $\mathcal{T} \vdash \neg Q$, the Soundness Theorem would imply that Q and $\neg Q$ must both be true, which gives a contradiction.
- In other words, if $\mathcal{T}$ has a model, then $\mathcal{T}$ is consistent.

Containing $\mathcal{PA}$.

- A theory is a set of formulas called axioms.
- $\mathcal{PA}$ is a specific theory whose formulas are the Peano axioms, which model the natural numbers $\mathbb{N}$.
- The hypothesis “ $\mathcal{T}$ contains $\mathcal{PA}$ ” means that all the Peano axioms must be axioms of $\mathcal{T}$.
- The theory $\mathcal{T}$ is then said to be “sufficiently rich,” meaning that we must be able to work with the natural numbers. This is a relatively weak requirement!

Recursive Axioms

- The hypothesis that $\mathcal{T}$ has recursive axioms (the usual term is “recursively axiomatizable”) is more technical. It means that we must be able to recognize the axioms.
- More precisely, for a given formula, we must be able to decide by a computation (or an algorithm) whether or not it is an axiom of $\mathcal{T}$.
- Once again, this is a reasonable hypothesis: to decide which theorems can be proved, we must be able to identify the axioms used as assumptions!

1.4. Second Theorem

Under the same assumptions as the First Incompleteness Theorem, we have the following result.

Theorem 3.

A consistent theory cannot prove its own consistency.

In concrete terms, this means that we cannot hope to verify the hypotheses of the Incompleteness Theorem while remaining within the theory $\mathcal{T}$; we would have to work in a broader framework.

For example, $\mathcal{PA}$ is generally considered to be a theory that has a model. In reality, we have implicitly accepted the existence of the natural numbers $\mathbb{N}$. To construct $\mathbb{N}$ rigorously, we must work in a broader framework, for example the theory $\mathcal{ZFC}$ (see the chapter “Peano Arithmetic”).

2. Consequences

2.1. Mathematical Consequences

Around 1900, Hilbert wanted to define a rigorous framework for the foundations of logic. During the 1800s, many mathematical concepts had already been formalized: the real numbers, limits, continuity, ... It was time to go further and completely axiomatize mathematics.

The goal was to abstract away from the physical world (continuity means drawing the graph without lifting your pencil) and write only logical formulas. However, problems quickly arose, such as Russell's paradox: it is not easy to define what a set is.

Gödel's First and Second Incompleteness Theorems shattered Hilbert's program. After Gödel, there was no longer any hope of obtaining a perfect axiomatization:

- on the one hand, if we work in a theory (sufficiently rich and consistent), there will be true but unprovable formulas,
- moreover, we cannot even prove that the theory is consistent in order to guarantee that we cannot prove both a statement and its opposite.

This was therefore fatal to Hilbert's program.

Despite this shock, Gödel's theorems had profound positive consequences extending far beyond a result concerning a small circle of logicians.

First, the tools developed by Gödel were forerunners of computer science. For example, the arithmetization of syntax (encoding every formula as a number) is a natural idea in the digital world in which we live today, but that was not the case in 1930.

More generally, Gödel was interested in determining what is computable, well before the notion of an algorithm and Turing's work (1936) on a theoretical computer, the Turing machine, and undecidability.

The Continuum Hypothesis.

Even though Gödel constructs an explicit formula $\mathcal{G}$ that is not provable, the only purpose of this formula is to prove the theorem.

The *continuum hypothesis* is a much more concrete statement that is also undecidable (neither $\mathcal{CH}$ nor $\neg\mathcal{CH}$ is provable in $\mathcal{ZF}$).

The independence result was completed by Cohen in 1963, thereby resolving the first of Hilbert's 23 problems.

Let us explain this problem.

- The sets $\mathbb{N}$, $\mathbb{Z}$, and $\mathbb{Q}$ are countable (in bijection with $\mathbb{N}$), so they have the same cardinality, even though they are infinite sets. We denote this cardinality by $\aleph_0$ (*aleph* is the first letter of the Hebrew alphabet).

- The sets $\mathbb{R}$, $[0, 1]$, $\mathbb{R}^2$, and $\mathbb{C}$ are uncountable; we denote their cardinality by $2^{\aleph_0}$ (an explanation of this notation follows below).
- The continuum hypothesis states that there is no other type of infinity between $\aleph_0$ and $2^{\aleph_0}$.
- In other words, if $\aleph_1$ denotes the first type of infinity strictly greater than $\aleph_0$, then $\aleph_1 = 2^{\aleph_0}$.
- Cohen thus proves that this question cannot be answered, which means that we can choose to accept or reject $\mathcal{CH}$ independently of everything else.

Let us return to the notation $2^{\aleph_0}$ for the cardinality of $\mathbb{R}$.

For two sets E, F , we sometimes denote by F^E the set of functions $f : E \rightarrow F$. This is justified by the fact that if E and F are finite, the number of such functions is $\text{Card}(F)^{\text{Card}(E)}$. For example, if $F = \{0, 1\}$ and $\text{Card}(E) = n$, then the number of functions $f : E \rightarrow \{0, 1\}$ is 2^n .

Let us explain how the real numbers in $[0, 1]$ can be described by functions $\mathbb{N} \rightarrow \{0, 1\}$. For $x \in [0, 1]$, we compute its binary expansion: $x = 0.100110101 \dots = \sum_{n=1}^{\infty} a_n 2^{-n}$, where $a_n \in \{0, 1\}$. This is like decimal notation, but with only the digits 0 and 1. Thus, with each binary expansion, we can associate a function $f : \mathbb{N} \rightarrow \{0, 1\}$ defined by $f(n) = a_n$. This correspondence is (roughly) bijective.

Thus, the cardinality of $[0, 1]$ is that of $\{0, 1\}^{\mathbb{N}}$, which is $2^{\aleph_0}$ (where $\aleph_0$ denotes the cardinality of $\mathbb{N}$).

This is also an opportunity to recall Cantor's theorem: $\text{Card}(\mathcal{P}(E)) > \text{Card}(E)$ for every set E . It is easy to see that $\mathcal{P}(\mathbb{N})$ corresponds to $\{0, 1\}^{\mathbb{N}}$. Indeed, $A \subset \mathbb{N}$ naturally defines a real number $x = 0.0110001 \dots$, where there is a 1 in the n th position exactly when $n \in A$. Thus, $[0, 1]$ has the same cardinality as $\mathcal{P}(\mathbb{N})$: $2^{\aleph_0}$, and this cardinality is strictly greater than $\aleph_0$.

2.2. Philosophical Consequence

The Incompleteness Theorems are among the few modern mathematical results known to nonmathematicians because of their influence on certain schools of thought.

These theorems make the fundamental difference between what is true and what is provable undeniable. This strengthens the Platonists' point of view: truth is an absolute reality, independent of the world around us.

Gödel's theorems also prove that mathematics is not merely a game of writing with abstract symbols, by showing that syntax (what can be proved) does not necessarily correspond to all of semantics (what is true).

They also raise questions about the difference between humans and machines. If we model a machine by a theory (with axioms corresponding to instructions), the machine will not be able to step outside its framework to prove its own consistency. For example, for Gödel's sentence $\mathcal{G}$, which says "I am not provable in $\mathcal{PA}$," we must work in a broader framework, for example $\mathcal{ZFC}$, to be able to say that this sentence is true. Can a machine encompass the full complexity of human reasoning?

2.3. What the Incompleteness Theorems Are Not

Completeness and incompleteness.

The Completeness Theorem and the Incompleteness Theorems are true theorems, proved by Gödel, and they do not contradict each other.

Recall completeness, for every formula $\mathcal{Q}$:

$$\mathcal{T} \models \mathcal{Q} \quad \text{iff} \quad \mathcal{T} \vdash \mathcal{Q}$$

Let us now work in the setting of a theory $\mathcal{T} = \mathcal{PA}$ (or one containing $\mathcal{PA}$) with $\mathbb{N}$ as a model.

The Incompleteness Theorem states that there is a formula $\mathcal{Q}$ such that:

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

It concerns specific theories (modeling $\mathbb{N}$), whereas completeness holds regardless of the structure $\mathcal{S}$. Incompleteness combined with completeness proves that $\mathcal{Q}$ cannot be true in every structure (because otherwise, by completeness, $\mathcal{Q}$ would be provable).

There are therefore structures $\mathcal{S}$, different from $\mathbb{N}$, in which $\mathcal{Q}$ is false (so we do not have $\mathcal{T} \models \mathcal{Q}$), even though $\mathcal{S}$ satisfies the Peano axioms. These are called nonstandard models of the natural numbers.

Are Mathematics and Logic Inconsistent?

The Second Theorem says that a theory cannot prove its own consistency. This means that, while remaining strictly within that theory, we can never be sure that it is impossible to prove both a formula $\mathcal{Q}$ and its negation $\neg\mathcal{Q}$.

However, if we work in a broader framework, we can manage. For example, the consistency of arithmetic can be proved from the theory $\mathcal{ZFC}$. An infinite-regress problem then arises: we cannot be sure of the consistency of $\mathcal{ZFC}$. We would have to work in an even larger theory...

There is therefore no single universal logical framework at the foundation of mathematics. The empirical consensus in modern mathematics is to use the framework $\mathcal{ZFC}$, in which no one has found an inconsistency.

3. Road Map

Proving the First Incompleteness Theorem is a major undertaking, quite technical but with several brilliant ideas (for example, the β function, which encodes an arbitrary list of numbers using only two integers!). Even one hundred years later, it remains difficult to approach. One can only admire the fact that a single person accomplished all of this.

We introduce the upcoming chapters one by one.

3.1. Peano Arithmetic

We describe the seven Peano axioms and the induction axiom schema that govern the behavior of the natural numbers. We also describe other, weaker theories (Presburger arithmetic and Robinson arithmetic) and a stronger theory, Zermelo–Fraenkel set theory, possibly augmented by the axiom of choice, $\mathcal{ZFC}$.

3.2. Gödel Coding

This is the “arithmetization of syntax” part: every formula, and even every proof, must be encoded by a single integer. We will then be able to give an informal proof that remains in the spirit of Gödel’s First Theorem, using a diagonalization argument. We will also take the opportunity to motivate this diagonalization method through Russell’s paradox and Cantor’s diagonal argument.

3.3. Primitive Recursive Functions

This is one of the technical chapters. Primitive recursive functions are the functions $f : \mathbb{N}^p \rightarrow \mathbb{N}$ that are “easily computable.” The goal is to see that operations on formulas, for example checking that a sequence of formulas is indeed a proof, are easily computable.

We will discuss the construction of Hilbert-style proofs, which are better suited than natural-deduction proofs to algorithmic treatment.

3.4. Representable Functions

This is another technical chapter, making the transition between the mathematical universe of the natural numbers and the language $\mathcal{L}_0$ of Peano arithmetic. We will prove that every primitive recursive function can be translated into a formula of $\mathcal{L}_0$. We thus obtain a formula that decides whether or not a proof is valid, as well as a formula expressing whether or not a statement is provable. It is in this chapter that we explain Gödel's β function.

3.5. Diagonalization

We explain how to obtain a formula $\mathcal{G}$ that speaks about itself.

We also discuss the liar paradox and a theorem of Tarski stating that there is no formula that defines the true formulas (whereas we have seen that we can detect provable formulas).

3.6. Proof of the Incompleteness Theorem

We state and prove the First Incompleteness Theorem, first in the setting of Peano arithmetic and then in the setting of broader theories. The proof is based on the diagonal lemma, which allows us to obtain a sentence that says, "I am not provable."

3.7. The Second Incompleteness Theorem

We conclude by stating the Second Theorem and proving it without going into detail.