

The Second Incompleteness Theorem

We conclude this book with Gödel's Second Incompleteness Theorem.

1. Gödel's Second Theorem

1.1. Statement

Theorem 1.

Let $\mathcal{T}$ be a consistent theory containing $\mathcal{PA}$ and having recursive axioms. Then $\mathcal{T}$ cannot prove its own consistency.

For a fixed theory $\mathcal{T}$, let $\text{Is-consistent}(\mathcal{T})$ denote the sentence expressing that $\mathcal{T}$ is consistent. The conclusion of the theorem is therefore:

$$\mathcal{T} \not\vdash \text{Is-consistent}(\mathcal{T})$$

Notice that the assumptions are the same as those of the First Incompleteness Theorem. In fact, we will prove the Second Incompleteness Theorem from the first, assuming its formalized version.

1.2. Consistency

Recall that a theory is *inconsistent* if it proves every formula, in particular the absurd formula " $0 = 1$ ". This can be expressed in the language $\mathcal{L}_0$ by $\text{Is-provable}(\ulcorner 0 = 1 \urcorner)$, where of course we write $1 = S(0)$. We therefore define the formula:

$$\text{Is-consistent}(\mathcal{T}) : \neg \text{Is-provable}(\ulcorner 0 = 1 \urcorner)$$

Thus the consistency of $\mathcal{T}$ can be expressed by a sentence of $\mathcal{L}_0$.

2. Proof

The proof follows fairly directly from the First Incompleteness Theorem if we accept its formalized version.

2.1. Formalization of the First Incompleteness Theorem

Suppose that $\mathcal{T}$ is a theory containing $\mathcal{PA}$ and that the axioms of $\mathcal{T}$ are recursive. We focus only on the consistency assumption.

Recall that the diagonal lemma provides a formula $\mathcal{G}$ such that:

$$\mathcal{T} \vdash (\mathcal{G} \leftrightarrow \neg \text{Is-provable}(\ulcorner \mathcal{G} \urcorner))$$

The statement of the First Incompleteness Theorem is then:

If $\mathcal{T}$ is consistent, then $\mathcal{G}$ is unprovable.

But by the diagonal lemma, $\mathcal{G}$ means exactly “ $\mathcal{G}$ is unprovable,” so the statement of the first theorem is also:

If $\mathcal{T}$ is consistent, then $\mathcal{G}$.

Here is the formalization of the First Incompleteness Theorem:

$$\mathcal{T} \vdash (\text{Is-consistent}(\mathcal{T}) \rightarrow \mathcal{G})$$

This is exactly the preceding statement, now *internalized*. The statement “if $\mathcal{T}$ is consistent, then $\mathcal{G}$ ” is provable (as we saw in the preceding chapters), but here we assume that a proof of it could be written within the theory $\mathcal{T}$. We will return to this formalization in the last section.

2.2. Proof of the Second Incompleteness Theorem

We assume the formalized version of the First Incompleteness Theorem in order to prove the second. In doing so, we follow the same approach as Gödel when he states the second theorem and sketches its proof at the end of his foundational paper on the Incompleteness Theorems.

We also need to assume the principle called “proof necessitation”:

$$\text{if } \mathcal{T} \vdash \mathcal{Q} \text{ then } \mathcal{T} \vdash \text{Is-provable}(\ulcorner \mathcal{Q} \urcorner)$$

This means that if a theory proves a formula, then it knows that it can prove that formula!

Let us now prove that a consistent theory cannot prove its own consistency.

We start with an arbitrary theory $\mathcal{T}$ and suppose that it proves its own consistency:

$$\mathcal{T} \vdash \text{Is-consistent}(\mathcal{T})$$

By *modus ponens* applied to the formalization of the First Incompleteness Theorem, we then have:

$$\begin{aligned} &\mathcal{T} \vdash \mathcal{G} \\ \text{therefore } &\mathcal{T} \vdash \text{Is-provable}(\ulcorner \mathcal{G} \urcorner) \text{ by proof necessitation} \\ \text{therefore } &\mathcal{T} \vdash \neg \mathcal{G} \text{ by the diagonal equivalence} \end{aligned}$$

In summary:

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

Therefore $\mathcal{T}$ is inconsistent.

Thus, by contraposition, we have proved the Second Incompleteness Theorem:

$$\text{If } \mathcal{T} \text{ is consistent, then } \mathcal{T} \nvdash \text{Is-consistent}(\mathcal{T}).$$

2.3. Returning to the formalization

The proof explained above is Gödel's original sketch (1931). A detailed proof would come only with the work of Hilbert and Bernays (1934–1939), followed by Löb (1955). This proof rests on three conditions, called the “Hilbert–Bernays–Löb derivability conditions”:

- **C1** - *Proof necessitation*.

$$\text{If } \mathcal{T} \vdash Q, \text{ then } \mathcal{T} \vdash \text{Is-provable}(\ulcorner Q \urcorner)$$

- **C2** - *Modus ponens for proofs*.

$$\mathcal{T} \vdash \text{Is-provable}(\ulcorner \mathcal{P} \rightarrow Q \urcorner) \rightarrow (\text{Is-provable}(\ulcorner \mathcal{P} \urcorner) \rightarrow \text{Is-provable}(\ulcorner Q \urcorner))$$

- **C3** - *Proof introspection*.

$$\mathcal{T} \vdash \text{Is-provable}(\ulcorner Q \urcorner) \rightarrow \text{Is-provable}(\ulcorner \text{Is-provable}(\ulcorner Q \urcorner) \urcorner)$$

We assume that these conditions hold. (They are highly technical to prove; it would take until 1955 to complete their formalization.) We can now give a rigorous proof of the formalization of the First Incompleteness Theorem. We start from the equivalence provided by diagonalization:

$$\mathcal{T} \vdash \mathcal{G} \leftrightarrow \neg \text{Is-provable}(\ulcorner \mathcal{G} \urcorner)$$

therefore $\mathcal{T} \vdash \mathcal{G} \rightarrow \neg \text{Is-provable}(\ulcorner \mathcal{G} \urcorner)$ keeping only the forward implication

therefore $\mathcal{T} \vdash \text{Is-provable}(\ulcorner \mathcal{G} \rightarrow \neg \text{Is-provable}(\ulcorner \mathcal{G} \urcorner) \urcorner)$ by (C1)

therefore $\mathcal{T} \vdash \text{Is-provable}(\ulcorner \mathcal{G} \urcorner) \rightarrow \text{Is-provable}(\ulcorner \neg \text{Is-provable}(\ulcorner \mathcal{G} \urcorner) \urcorner)$ by (C2)

But by (C3), we always have:

$$\mathcal{T} \vdash \text{Is-provable}(\ulcorner \mathcal{G} \urcorner) \rightarrow \text{Is-provable}(\ulcorner \text{Is-provable}(\ulcorner \mathcal{G} \urcorner) \urcorner)$$

Setting

$$Q : \text{Is-provable}(\ulcorner \mathcal{G} \urcorner),$$

we obtain two conclusions:

$$\mathcal{T} \vdash \text{Is-provable}(\ulcorner \mathcal{G} \urcorner) \rightarrow \text{Is-provable}(\ulcorner Q \urcorner) \wedge \text{Is-provable}(\ulcorner \neg Q \urcorner)$$

therefore $\mathcal{T} \vdash \text{Is-provable}(\ulcorner \mathcal{G} \urcorner) \rightarrow \text{Is-provable}(\ulcorner Q \wedge \neg Q \urcorner)$

therefore $\mathcal{T} \vdash \text{Is-provable}(\ulcorner \mathcal{G} \urcorner) \rightarrow \text{Is-provable}(\ulcorner 0 = 1 \urcorner)$

thus $\mathcal{T} \vdash \text{Is-provable}(\ulcorner \mathcal{G} \urcorner) \rightarrow \neg \text{Is-consistent}(\mathcal{T})$

By contraposition:

$$\mathcal{T} \vdash \text{Is-consistent}(\mathcal{T}) \rightarrow \neg \text{Is-provable}(\ulcorner \mathcal{G} \urcorner)$$

But by diagonalization, the conclusion is exactly $\mathcal{G}$, and therefore:

$$\mathcal{T} \vdash \text{Is-consistent}(\mathcal{T}) \rightarrow \mathcal{G}$$

This is the desired formalization of the First Incompleteness Theorem.