inp.1 The Second Incompleteness Theorem

inc:inp:2in:sec

How can we express the assertion that **PA** doesn't prove its own consistency? Saying **PA** is inconsistent amounts to saying that $\mathbf{PA} \vdash 0 = 1$. So we can take the consistency statement $\mathsf{Con}_{\mathbf{PA}}$ to be the sentence $\lnot \mathsf{Prov}_{\mathbf{PA}}(\ulcorner 0 = 1 \urcorner)$, and then the following theorem does the job:

inc:inp:2in:thm:second-incompleteness

Theorem inp.1. *Assuming* **PA** *is consistent, then* **PA** *does not derive* $\mathsf{Con}_{\mathbf{PA}}$.

It is important to note that the theorem depends on the particular representation of $\mathsf{Con}_{\mathbf{PA}}$ (i.e., the particular representation of $\mathsf{Prov}_{\mathbf{PA}}(y)$). All we will use is that the representation of $\mathsf{Prov}_{\mathbf{PA}}(y)$ satisfies the three derivability conditions, so the theorem generalizes to any theory with a derivability predicate having these properties.

It is informative to read Gödel's sketch of an argument, since the theorem follows like a good punch line. It goes like this. Let $\gamma_{\mathbf{PA}}$ be the Gödel sentence that we constructed in the proof of **??**. We have shown "If **PA** is consistent, then **PA** does not derive $\gamma_{\mathbf{PA}}$." If we formalize this *in* **PA**, we have a proof of

$$\mathsf{Con}_{\mathbf{PA}} \to \lnot \mathsf{Prov}_{\mathbf{PA}}(\ulcorner \gamma_{\mathbf{PA}} \urcorner).$$

Now suppose **PA** derives $\mathsf{Con}_{\mathbf{PA}}$. Then it derives $\lnot \mathrm{Prov}_{\mathbf{PA}}(\ulcorner \gamma_{\mathbf{PA}} \urcorner)$. But since $\gamma_{\mathbf{PA}}$ is a Gödel sentence, this is equivalent to $\gamma_{\mathbf{PA}}$. So **PA** derives $\gamma_{\mathbf{PA}}$.

But: we know that if **PA** is consistent, it doesn't derive $\gamma_{\mathbf{PA}}$! So if **PA** is consistent, it can't derive $\mathsf{Con}_{\mathbf{PA}}$.

To make the argument more precise, we will let $\gamma_{\mathbf{PA}}$ be the Gödel sentence for **PA** and use the derivability conditions (P1)–(P3) to show that **PA** derives $\mathsf{Con}_{\mathbf{PA}} \to \gamma_{\mathbf{PA}}$. This will show that **PA** doesn't derive $\mathsf{Con}_{\mathbf{PA}}$. Here is a sketch of the proof, in **PA**. (For simplicity, we drop the **PA** subscripts.)

$$\gamma \leftrightarrow \neg \mathsf{Prov}(\ulcorner \gamma \urcorner) \tag{1}$$ inc:inp:2in:G2-1

γ is a Gödel sentence

$$\gamma \to \neg \mathsf{Prov}(\ulcorner \gamma \urcorner) \tag{2}$$ inc:inp:2in:G2-2

from eq. (1)

$$\gamma \to (\mathsf{Prov}(\ulcorner \gamma \urcorner) \to \bot) \tag{3}$$ inc:inp:2in:G2-3

from eq. (2) by logic

$$\mathsf{Prov}(\ulcorner \gamma \to (\mathsf{Prov}(\ulcorner \gamma \urcorner) \to \bot) \urcorner) \tag{4}$$ inc:inp:2in:G2-4

by from eq. (3) by condition P1

$$\mathsf{Prov}(\ulcorner \gamma \urcorner) \to \mathsf{Prov}(\ulcorner (\mathsf{Prov}(\ulcorner \gamma \urcorner) \to \bot) \urcorner) \tag{5}$$ inc:inp:2in:G2-5

from eq. (4) by condition P2

$$\mathsf{Prov}(\ulcorner \gamma \urcorner) \to (\mathsf{Prov}(\ulcorner \mathsf{Prov}(\ulcorner \gamma \urcorner) \urcorner) \to \mathsf{Prov}(\ulcorner \bot \urcorner)) \tag{6}$$ inc:inp:2in:G2-6

from eq. (5) by condition P2 and logic

$$\mathsf{Prov}(\ulcorner \gamma \urcorner) \to \mathsf{Prov}(\ulcorner \mathsf{Prov}(\ulcorner \gamma \urcorner) \urcorner) \tag{7}$$ inc:inp:2in:G2-7

by P3

$$\mathsf{Prov}(\ulcorner \gamma \urcorner) \to \mathsf{Prov}(\ulcorner \bot \urcorner) \tag{8}$$ inc:inp:2in:G2-8

from eq. (6) and eq. (7) by logic

$$\mathsf{Con} \to \neg \mathsf{Prov}(\ulcorner \gamma \urcorner) \tag{9}$$ inc:inp:2in:G2-9

contraposition of eq. (8) and $\mathsf{Con} \equiv \neg \mathsf{Prov}(\ulcorner \bot \urcorner)$

$$\mathsf{Con} \to \gamma$$

from eq. (1) and eq. (9) by logic

The use of logic in the above just elementary facts from propositional logic, e.g., eq. (3) uses $\vdash \neg\varphi \leftrightarrow (\varphi \to \bot)$ and eq. (8) uses $\varphi \to (\psi \to \chi), \varphi \to \psi \vdash \varphi \to \chi$. The use of condition P2 in eq. (5) and eq. (6) relies on instances of P2, $\mathsf{Prov}(\ulcorner \varphi \to \psi \urcorner) \to (\mathsf{Prov}(\ulcorner \varphi \urcorner) \to \mathsf{Prov}(\ulcorner \psi \urcorner))$. In the first one, $\varphi \equiv \gamma$ and $\psi \equiv \mathsf{Prov}(\ulcorner \gamma \urcorner) \to \bot$; in the second, $\varphi \equiv \mathsf{Prov}(\ulcorner G \urcorner)$ and $\psi \equiv \bot$.

The more abstract version of the second incompleteness theorem is as follows:

Theorem inp.2. *Let* **T** *be any consistent, axiomatized theory extending* **Q** *and let* $\mathsf{Prov}_T(y)$ *be any formula satisfying derivability conditions P1–P3 for* **T**. *Then* **T** *does not derive* Con_T. *inc:inp:2in:thm:second-incompleteness-gen*

Problem inp.1. Show that **PA** derives $\gamma_{\mathbf{PA}} \to \mathsf{Con}_{\mathbf{PA}}$.

digression The moral of the story is that no "reasonable" consistent theory for mathematics can derive its own consistency statement. Suppose **T** is a theory of mathematics that includes **Q** and Hilbert's "finitary" reasoning (whatever that may be). Then, the whole of **T** cannot derive the consistency statement of **T**, and so, a fortiori, the finitary fragment can't derive the consistency statement of **T** either. In that sense, there cannot be a finitary consistency proof for "all of mathematics."

There is some leeway in interpreting the term "finitary," and Gödel, in the 1931 paper, grants the possibility that something we may consider "finitary" may lie outside the kinds of mathematics Hilbert wanted to formalize. But Gödel was being charitable; today, it is hard to see how we might find something that can reasonably be called finitary but is not formalizable in, say, **ZFC**.

Photo Credits

Bibliography