Gödel Without Tears – 10

Introducing the Second Theorem

Peter Smith

January 22, 2010

- Con_T, a canonical consistency sentence for T
- The formalized First Theorem
- The Second Theorem
- Why it matters
- What it takes to prove it

This episode introduces the Second Incompleteness Theorem, says something about what it takes to prove it, and why it matters.

Just two very quick reminders before we start. We said

Defn. 50. $\mathsf{Prf}_T(\mathsf{x}, \mathsf{y})$ *stands in for a T-wff that canonically captures Prf_T.*

Defn. 51. *Put* $\mathsf{Prov}_T(\mathsf{y}) =_{\mathrm{def}} \exists \mathsf{v}\, \mathsf{Prf}_T(\mathsf{v}, \mathsf{y})$*: we'll call such an expression* a provability predicate *for T.*

And then recall we proved:

Theorem 47. *If T is p.r. axiomatized and contains* Q*,* $T \vdash \mathsf{G}_T \leftrightarrow \neg\mathsf{Prov}_T(\overline{\ulcorner \mathsf{G_T} \urcorner})$.

38 The Second Theorem introduced

38.1 Definitional preliminaries

We haven't put any requirement on the particular formulation of first-order logic built into Q (and hence any theory which contains it). It may or may not have a built-in absurdity constant. But henceforth,

Defn. 58. $\perp$ *is T's built-in absurdity constant if it has one, or else can be treated as abbreviating* $\mathsf{0} = \overline{\mathsf{1}}$.

If T contains Q, then on either reading of $\perp$, T is consistent if and only if it *doesn't* prove $\perp$. That motivates the definition

Defn. 59. Con_T *abbreviates* $\neg\mathsf{Prov}_T(\overline{\ulcorner \perp \urcorner})$.

Note that since Prov_T is Σ_1, Con_T is Π_1. For obvious reasons, the arithmetic sentence Con_T is called a canonical consistency sentence for T, and is true if and only if T is consistent.

Or at least, that's the crispest definition of a consistency sentence for T. There are alternatives. Here's another natural one. Suppose $\mathsf{Contr}(\mathsf{x}, \mathsf{y})$ captures the p.r. relation which holds between two numbers when one codes for some sentence φ and the other for $\neg\varphi$. Then we could put

Defn. 60. $\mathsf{Con}'_T =_{\mathrm{def}} \neg\exists \mathsf{x} \exists \mathsf{y}(\mathsf{Prov}_T(\mathsf{x}) \wedge \mathsf{Prov}_T(\mathsf{y}) \wedge \mathsf{Contr}(\mathsf{x}, \mathsf{y}))$.

But, on modest assumptions, this sort of definition and its variants are equivalent: so we'll stick to the crisp one.

38.2 The unprovability of consistency

One half of the First Theorem tells us that, for nice enough T,

(1) If T is consistent then G_T is not provable in T.

Now, we can faithfully express (1) inside T by

(2) $\mathsf{Con}_T \to \neg\mathsf{Prov}_T(\overline{\ulcorner \mathsf{G}_T \urcorner})$.

But now reflect that the informal reasoning for the First Theorem is in fact rather elementary (we needed no higher mathematics at all, just simple reasoning about arithmetic matters). So we might expect that if T contains enough arithmetic, it should itself be able to replicate that elementary reasoning. So we have that a strong enough T can not only express (half) of the First Theorem, but can prove it too! – so

Theorem 53. *For strong enough T, $T \vdash \mathsf{Con}_T \to \neg\mathsf{Prov}_T(\overline{\ulcorner \mathsf{G}_T \urcorner})$.*

Call such a result the *Formalized First Theorem* for the relevant provability predicate.

Now, we've just reminded ourselves of Theorem 51 which says that $T \vdash \mathsf{G}_T \leftrightarrow \neg\mathsf{Prov}_T(\overline{\ulcorner \mathsf{G}_T \urcorner})$. So putting those facts together, we get

(3) if $T \vdash \mathsf{Con}_T$ then $T \vdash \mathsf{G}_T$.

And we know from the First Theorem that,

(4) If T is consistent, $T \nvdash \mathsf{G}_T$.

So the Formalized First Theorem immediately yields the unprovability of the relevant consistency sentence.

Theorem 54. *For strong enough T, if T is consistent, then $T \nvdash \mathsf{Con}_T$.*

Which is a vague version Second Incompleteness Theorem: roughly, for the right kind of theories T and the right kind of consistency sentences Con_T, T can't prove its own consistency sentences.

Obviously, we'll need in say something about 'for strong enough T'; but this will do as an introduction. Indeed, this is about as much as Gödel says in his original 1931 paper where he too didn't spell out the details.

39 How interesting is the Second Theorem?

You might well think: 'OK, so we can't (for example) derive Con_T in T. But that fact is of course no evidence at all *against* T's consistency, since we already know from the First Theorem that lots of true claims about provability are underivable in T. While if, *per impossibile*, we could have given a T proof of Con_T, that wouldn't have given us any special evidence *for* T's consistency – we could simply reflect that even if T were inconsistent we'd still be able to derive Con_T, since we can derive *anything* in an inconsistent theory! Hence the derivability or otherwise of a canonical statement of T's consistency inside T itself can't show us a great deal.'

But, on reflection, the Theorem *does* yield some plainly important and substantial corollaries, of which the most important is this:

Theorem 55. *Suppose S is a strong enough theory for the Second Theorem to apply to it, and W is a weaker fragment of S, then $W \nvdash \mathsf{Con}_S$.*

That's because, if S can't prove Con_S, a fortiori *part* of S can't prove it .

So, for example, we *can't* take some problematic rich theory like set theory which extends arithmetic and show that it is consistent by (i) using arithmetic coding for talking about its proofs and then (ii) using uncontentious reasoning already available in some relatively weak, purely arithmetical, theory which is a fragment of set theory.

Which means that the Second Theory – at least at first blush – sabotages Hilbert's Programme (see §24.3).

40 What does it take to prove the Second Theorem?

$\mathsf{Prov}_T(\mathsf{y})$ abbreviates $\exists \mathsf{v}\, \mathsf{Prf}_T(\mathsf{v},\mathsf{y})$; and Prf_T is a Σ expression. To prove the likes of

Theorem 53. *For strong enough T, $T \vdash \mathsf{Con}_T \to \neg\mathsf{Prov}_T(\overline{\ulcorner \mathsf{G}_T \urcorner})$.*

we need therefore to be handle quantified claims involving Σ expressions. And how do we prove quantified claims? Using induction is the default method.

It therefore seems quite a good bet that T with be able to prove (the relevant half of) the First Theorem for T only if T has Σ_1-induction – meaning that T's axioms include (the universal closures of) all instances of the first-order Induction Schema where the induction predicate φ is Σ_1. So let's define

Defn. 61. *A theory is Σ-normal, if it is p.r. axiomatized, contains* Q*, and also also includes induction at least for Σ_1 wffs.*

Then the following looks a plausible conjecture

Theorem 56. *If T is Σ-normal, then T proves the formalized First Theorem, and so $T \nvdash \mathsf{Con}_T$.*

Which is indeed provable.

That's all short and sweet, and is the basic headline news about the Second Theorem. But for more at this level, see *IGT*, Ch. 24.