12 The First Incompleteness Theorem, syntactic version

We now use the same construction of a Gödel sentence as in the previous chapter to show again that PA is incomplete, but this time making only syntactic assumptions. And then we show how to generalize this syntactic version of the incompleteness theorem.

12.1 ω-completeness, ω-consistency

We need to define two key notions. We'll assume in this section that we are dealing with theories whose language includes the language of basic arithmetic. And take all the quantifiers mentioned to run over the natural numbers.[1]

First, then,

Defn. 49. *A theory T is* ω-incomplete *iff, for some open wff $\varphi(\mathsf{x})$, T can prove $\varphi(\overline{\mathsf{n}})$ for each natural number n, but T can't go on to prove $\forall \mathsf{x}\varphi(\mathsf{x})$.*

We saw in §5.6 that Q is ω-incomplete: that's because it can prove each instance of $\mathsf{0} + \overline{\mathsf{n}} = \overline{\mathsf{n}}$, but can't prove $\forall \mathsf{x}(\mathsf{0} + \mathsf{x} = \mathsf{x})$. We added induction to Q hoping to repair as much ω-incompleteness as we could: but, as we'll see, PA remains ω-incomplete, assuming it is consistent.[2]

Second, we want the following idea:

Defn. 50. *A theory T is* ω-inconsistent *iff, for some open wff $\varphi(\mathsf{x})$, T can prove each $\varphi(\overline{\mathsf{n}})$ and T can also prove $\neg\forall \mathsf{x}\varphi(\mathsf{x})$.*

Or, entirely equivalently of course, we could say that T is ω-inconsistent if, for some open wff $\psi(\mathsf{x})$, $T \vdash \exists \mathsf{x}\psi(\mathsf{x})$, yet for each number n we have $T \vdash \neg\psi(\overline{\mathsf{n}})$.

Note that ω-inconsistency, like ordinary inconsistency, is a syntactically defined property: it is characterized in terms of what wffs can be proved by the theory, not in terms of what the wffs mean. Note too that, in a classical context,

---

[1] If necessary, therefore, read $\forall \mathsf{x}\varphi(\mathsf{x})$ as a restricted quantifier $\forall \mathsf{x}(\mathsf{N}\mathsf{x} \to \varphi(\mathsf{x}))$, where '$\mathsf{N}$' picks out the numbers from the domain of the theory's native quantifiers (see Defn. 11).

[2] Why the 'ω' in 'ω-incomplete'? Because 'ω' is a standard label for the set of natural numbers (when equipped with their usual ordering).

ω-consistency – defined of course as not being ω-inconsistent! – trivially implies plain consistency. That's because T's being ω-consistent is a matter of its *not* being able to prove a certain combination of wffs, which entails that T can't prove *all* wffs, hence T can't be inconsistent.

Now compare and contrast. Suppose T can prove $\varphi(\overline{\mathsf{n}})$ for each n. T is ω-incomplete if it can't prove something we'd then also like it to prove, namely $\forall \mathsf{x}\varphi(\mathsf{x})$. While T is ω-inconsistent if it can actually prove the *negation* of what we'd like it to prove, i.e. it can prove $\neg\forall \mathsf{x}\varphi(\mathsf{x})$.

So ω-incompleteness in a theory of arithmetic is a regrettable weakness. But ω-inconsistency is a Very Bad Thing (not as bad as outright inconsistency, maybe, but still bad enough). For evidently, a theory T that can prove each of $\varphi(\overline{\mathsf{n}})$ and yet also prove $\neg\forall \mathsf{x}\varphi(\mathsf{x})$ is just not going to be an acceptable candidate for regimenting arithmetic.

Bring semantic ideas back into play for a moment. Suppose T's standard numerals denote the numbers and the quantifier here runs over the natural numbers. Then it can't be the case that each of $\varphi(\overline{\mathsf{n}})$ is true and yet $\neg\forall \mathsf{x}\varphi(\mathsf{x})$ is true too. So our ω-inconsistent T can't be sound.

Given that we want formal arithmetics to have theorems which *are* all true on a standard interpretation, we must therefore want ω-consistent arithmetics. And given that we think e.g. PA *is* sound on its standard interpretation, we are committed to thinking that it *is* ω-consistent.

12.2 The First Theorem for PA – the syntactic version

G is by definition the diagonalization of the open wff $\mathsf{U} =_{\text{def}} \forall \mathsf{x}\neg\mathsf{Gdl}(\mathsf{x},\mathsf{y})$, i.e. G is the wff $\forall \mathsf{x}\neg\mathsf{Gdl}(\mathsf{x},\overline{\ulcorner \mathsf{U} \urcorner})$ (see §11.1). And now recall Defn. 46: the wff Gdl by hypothesis doesn't just express *Gdl* but *captures* it. Using this fact about Gdl, we can again show that PA does not prove G, but this time *without* making the semantic assumption that PA is sound:

Theorem 39. *If* PA *is consistent,* $\mathsf{PA} \nvdash \mathsf{G}$.

Proof. Suppose that $\mathsf{PA} \vdash \mathsf{G}$.

If G has a proof, then there is some super g.n. m that codes its proof. But G is the diagonalization of the wff with g.n. $\ulcorner \mathsf{U} \urcorner$. Hence, by definition, $Gdl(m, \ulcorner \mathsf{U} \urcorner)$.

Since Gdl captures the relation *Gdl* and we have $Gdl(m, \ulcorner \mathsf{U} \urcorner)$, then by the definition of capturing (i) $\mathsf{PA} \vdash \mathsf{Gdl}(\overline{\mathsf{m}}, \overline{\ulcorner \mathsf{U} \urcorner})$.

But our supposition $\mathsf{PA} \vdash \mathsf{G}$, i.e. $\mathsf{PA} \vdash \forall \mathsf{x}\neg\mathsf{Gdl}(\mathsf{x},\overline{\ulcorner \mathsf{U} \urcorner})$, implies (ii) $\mathsf{PA} \vdash \neg\mathsf{Gdl}(\overline{\mathsf{m}}, \overline{\ulcorner \mathsf{U} \urcorner})$, just by instantiating the quantifier.

So, combining (i) and (ii), the supposition that $\mathsf{PA} \vdash \mathsf{G}$ entails that PA is inconsistent.

Therefore, if PA is consistent, $\mathsf{PA} \nvdash \mathsf{G}$. □

We'll now show that PA also can't prove the *negation* of G, again without assuming PA's soundness:

Theorem 40. *If* PA *is ω-consistent,* $\mathsf{PA} \nvdash \neg\mathsf{G}$.

Proof. Suppose PA is ω-consistent and $\mathsf{PA} \vdash \neg\mathsf{G}$. We derive a contradiction, and the theorem follows.

Given PA is ω-consistent, it is consistent. So given PA proves $\neg\mathsf{G}$, it can't prove G. So no number m is the super g.n. of a proof for G. But, again, G is the diagonalization of the wff with g.n. $\ulcorner\mathsf{U}\urcorner$. Hence, for every number m, $Gdl(m, \ulcorner\mathsf{U}\urcorner)$ is false.

Since Gdl captures the relation Gdl and $Gdl(m, \ulcorner\mathsf{U}\urcorner)$ is false for each m, by the definition of capturing we have (i) for each m, $\mathsf{PA} \vdash \neg\mathsf{Gdl}(\overline{\mathsf{m}}, \overline{\ulcorner\mathsf{U}\urcorner})$.

But our supposition $\mathsf{PA} \vdash \neg\mathsf{G}$ is equivalent to (ii) $\mathsf{PA} \vdash \exists\mathsf{x}\mathsf{Gdl}(\mathsf{x}, \overline{\ulcorner\mathsf{U}\urcorner})$.

Combining (i) and (ii), PA is ω-inconsistent, contradicting our initial supposition. □

Now recall that G is a Π_1 sentence. That observation put together with what we've shown in this section gives us the following portmanteau result:

Theorem 41. *If* PA *is consistent, then there is a* Π_1 *sentence* G *such that* $\mathsf{PA} \nvdash \mathsf{G}$*, and if* PA *is ω-consistent* $\mathsf{PA} \nvdash \neg\mathsf{G}$*. Hence, assuming ω-consistency and so consistency,* PA *is negation incomplete.*

12.3 Two quick corollaries

Theorem 42. *If* PA *is consistent, it is ω-incomplete.*

Proof. Assume PA is consistent. By Theorem 39, it doesn't prove G.

That means that no number m is the super g.n. of a proof of G. So, exactly as in the proof of the previous theorem, (i) for each m, $\mathsf{PA} \vdash \neg\mathsf{Gdl}(\overline{\mathsf{m}}, \overline{\ulcorner\mathsf{U}\urcorner})$.

But since G is unprovable, (ii) $\mathsf{PA} \nvdash \forall\mathsf{x}\neg\mathsf{Gdl}(\mathsf{x}, \overline{\ulcorner\mathsf{U}\urcorner})$.

The combination of (i) and (ii) shows that PA is ω-incomplete. □

Theorem 43. *If* PA *is consistent, then* $\mathsf{PA} + \neg\mathsf{G}$ *(the theory you get by adding* $\neg\mathsf{G}$ *as an additional axiom) is also consistent but is ω-inconsistent.*

Proof. Assume PA is consistent. If $\mathsf{PA} + \neg\mathsf{G}$ were inconsistent, then PA would prove G, contrary to Theorem 39. So $\mathsf{PA} + \neg\mathsf{G}$ is also consistent.

Now, since the expanded theory can prove everything PA can prove, we know as before that (i) for each m, $\mathsf{PA} + \neg\mathsf{G} \vdash \neg\mathsf{Gdl}(\overline{\mathsf{m}}, \overline{\ulcorner\mathsf{U}\urcorner})$.

But just by the definition of $\neg\mathsf{G}$, (ii) $\mathsf{PA} + \neg\mathsf{G} \vdash \exists\mathsf{x}\mathsf{Gdl}(\mathsf{x}, \overline{\ulcorner\mathsf{U}\urcorner})$.

And (i) and (ii) together imply that $\mathsf{PA} + \neg\mathsf{G}$ is ω-inconsistent. □

12.4 Generalizing the proof

The proof for Theorem 41 evidently generalizes. Suppose T is a p.r. axiomatized theory which (perhaps after introducing some new vocabulary by definitions)

contains Q – in the obvious sense that the language of T includes the language of basic arithmetic, and T can prove every Q-theorem. Then, assuming we are working with normal scheme for Gödel-numbering wffs of T, the relation $Gdl_T(m, n)$ which holds when m numbers a T-proof of the diagonalization of the wff with number n will be primitive recursive again.

Since T can prove everything Q proves, T will be able to capture the p.r. relation Gld_T by a Σ_1 wff Gld_T. Just as we did for PA, we'll be able to construct the corresponding Π_1 wff G_T. And, exactly the same arguments as before will then show, more generally,

Theorem 44. *If T is a consistent p.r. axiomatized theory which contains* Q*, then there will be a Π_1 sentence G_T such that $T \nvdash \mathsf{G}_T$, and if T is ω-consistent, $T \nvdash \neg\mathsf{G}_T$. Hence, assuming ω-consistency and so consistency, T is negation incomplete.*

And note, this gives us another incompletability theorem: if we keep chucking more and more additional axioms at our theory T, it will still remain negation incomplete, unless it stops ω-consistent or stops being p.r. axiomatized.

When people refer to the *First Incompleteness Theorem* (without qualification), they typically mean something like our last Theorem, deriving incompleteness from *syntactic* assumptions. Let's re-emphasize that last point. Being p.r. axiomatized is a syntactic property; containing Q is a matter of Q's axioms again being adopted as axioms or being derivable, a syntactic property; being consistent here is a matter of no contradictory pair φ, $\neg\varphi$ being derivable; being ω-consistent is another syntactic property as we stressed before. The chains of argument that lead to our Theorem depend just on the given syntactic assumptions, via e.g. the proof that Q can capture all p.r. functions – another claim about a syntactically definable property. That is why we are calling this the *syntactic* incompleteness theorem.

Of course, we are *interested* in these various syntactically definable properties because of their semantic relevance: for example, we care about the idea of capturing p.r. functions because we are interested in what an interpreted theory might be able to prove in the sense of establish-as-true. But it is one thing for us to have a semantic motivation for being interested in a certain concept, it is another thing for that concept to have semantic content.

12.5 Comparing old and new syntactic incompleteness theorems

Compare Theorem 44 with our initially announced

Theorem 2. *Suppose T is a formal axiomatized theory whose language contains the language of basic arithmetic. Then, if T is consistent and can prove a certain modest amount of arithmetic (and has a certain additional property that any sensible formalized arithmetic will share), there will be a sentence G_T of basic arithmetic such that $T \nvdash \mathsf{G}_T$ and $T \nvdash \neg\mathsf{G}_T$, so T is negation incomplete.*

Our new theorem fills out the old one in various respects. It fixes the 'modest amount of arithmetic' that T is assumed to contain and it also spells out the 'additional desirable property' of ω-consistency which we previously left mysterious. Further it tells us more about the undecidable Gödel sentence – namely it has minimal quantifier complexity, i.e. it is a Π_1 sentence of arithmetic. Our new theorem is weaker, however, as it only applies to p.r. axiomatized theories, not to effectively axiomatized theories more generally. But we've already noted, that that's not much loss. (And again, if we insist, we can in fact go on to make up the shortfall: see the next Interlude.)

12.6 Gödel's own Theorem

As we said, Theorem 44, or something like it, is what people usually mean when they speak without qualification of 'The First Incompleteness Theorem'. But since the stated theorem refers to Robinson Arithmetic Q (developed by Robinson in 1950), and Gödel didn't originally know about that (in 1931), our version can't be quite what Gödel originally proved. But it is a near miss. Let's explain.

Looking again at our analysis of the syntactic argument for incompleteness, we see that we are interested in theories which extend Q *because we are interested in theories which can capture p.r. relations like Gdl*. It's being able to capture *Gdl* that is the crucial condition for our proof to go through. So let's say

Defn. 51. *A theory T is* p.r. adequate *if it can capture all primitive recursive functions and relations.*

Then, instead of mentioning Q, let's instead explicitly write in the requirement of p.r. adequacy. So, by just the same arguments,

Theorem 45. *If T is a p.r. adequate, p.r. axiomatized theory whose language includes L_A, then there is Π_1 sentence φ such that, if T is consistent then $T \nvdash \varphi$, and if T is ω-consistent then $T \nvdash \neg\varphi$.*

And *this* is pretty much Gödel's own general version of the incompleteness result. I suppose that it has as much historical right as any to be called *Gödel's* First Theorem.

('Hold on! If *that's* the original First Theorem, we don't need to do the hard work showing that Q and PA are p.r. adequate, do we?' Well, yes and no. No, proving *this* original version of the Theorem of course doesn't depend on proving that any particular theory is p.r. adequate. But yes, showing that this Theorem has real bite, showing that it does actually apply to familiar arithmetics, does depend on proving that these arithmetics are indeed p.r. adequate.)

Thus, in his 1931 paper, Gödel first proves his Theorem VI which, with a bit of help from his Theorem VIII shows that the formal system P – which is his

simplified version of the hierarchical type-theory of *Principia Mathematica* – has a formally undecidable Π_1 sentence.[3] Then he immediately generalizes:

> In the proof of Theorem VI no properties of the system P were used besides the following:
>
> 1. The class of axioms and the rules of inference (that is, the relation 'immediate consequence') are [primitive] recursively definable (as soon as we replace the primitive signs in some way by the natural numbers).
> 2. Every [primitive] recursive relation is definable [i.e., in our terms, is 'capturable'] in the system P.
>
> Therefore, in every formal system that satisfies the assumptions 1 and 2 and is ω-consistent, there are undecidable propositions of the form $[\forall x F(x)]$, where F is a [primitive] recursively defined property of natural numbers, and likewise in every extension of such a system by a recursively definable ω-consistent class of axioms.

Which gives us our Theorem 45.

[3] Or as Gödel put it, the undecidable sentence is 'of Goldbach type'. The allusion here is to Goldbach's conjecture that every even number other than two is the sum of two primes. The claim that the particular number n is an even number other than two and is the sum of two primes is expressible by a Δ_0 wff (why?). So Goldbach's conjecture, the universal quantification of this claim about n, is expressible by a Π_1 wff.

13 The Diagonalization Lemma, and Rosser's Theorem

In the previous chapter, we proved Theorem 44, a version of the syntactic First Theorem. In this chapter, we start by proving this incompleteness theorem again by a slightly different route – this time highlighting the so-called *Diagonalization Lemma*.

The same important Lemma can be used in proving further important theorems. We will look at two. The first of these is due to J. Barkley Rosser, and shows us how to tweak the Gödelian syntactic theorem so that we no longer need the assumption of ω-consistency. Then, in the short following chapter, we show how we can very easily derive an important two-part theorem about truth usually ascribed to Alfred Tarski (though known to Gödel).

13.1 The Diagonalization Lemma

First, it's useful to recall the definition of capturing a function:

Defn. 37. *The theory T captures the one-place function f by the open wff $\psi(\mathsf{x},\mathsf{y})$ iff, for any m, n,*

i. if $f(m) = n$, then $T \vdash \psi(\overline{\mathsf{m}}, \overline{\mathsf{n}})$,
ii. if $f(m) \neq n$, then $T \vdash \neg\psi(\overline{\mathsf{m}}, \overline{\mathsf{n}})$.
iii. $T \vdash \exists!\mathsf{y}\psi(\overline{\mathsf{m}}, \mathsf{y})$.

And this time, for use in just a moment, let's note explicitly that (i) and (iii) together imply

iv. if $f(m) = n$, then $T \vdash \forall\mathsf{x}(\psi(\overline{\mathsf{m}}, \mathsf{z}) \leftrightarrow \mathsf{z} = \overline{\mathsf{n}})$.

Next, let's restate Theorem 32 about the *diag* function, and apply Theorems 27 and 28 to add an important clause about expressing and capturing it (we assume as usual that we have normal Gödel-numbering scheme in place for the relevant theory T):

Theorem 46. *If T is a p.r. axiomatized theory which contains Q, there is a p.r. function $diag_T(n)$ which, when applied to a number n which is the g.n. of some T-wff with one free variable, yields the g.n. of that wff's diagonalization, and*

yields n otherwise. And, as with any p.r. function, T can express and capture this function by a Σ_1 *wff* $\mathsf{Diag}_T(\mathsf{x},\mathsf{y})$.

We can now officially state and prove the two-part *Diagonalization Lemma* (Rudolf Carnap noted a version of part (i): he is often, but I think wrongly, attributed (ii) as well):

Theorem 47. *If T is a p.r. axiomatized theory which contains* Q*, and* φ *is a one-place open sentence of T's language, then there is sentence* δ *such that (i)* $\delta \leftrightarrow \varphi(\overline{\ulcorner\delta\urcorner})$ *is true, and moreover (ii)* $T \vdash \delta \leftrightarrow \varphi(\overline{\ulcorner\delta\urcorner})$.

To avoid unsightly rashes of subscripts, let's drop subscript 'T's. Then we can argue as follows. (The proofs *look* complicated at first glance; but after the initial construction of δ, we are just applying various definitions and easy consequences.)

Proof for (i). Put $\alpha =_{\text{def}} \forall \mathsf{z}(\mathsf{Diag}(\mathsf{y},\mathsf{z}) \to \varphi(\mathsf{z}))$, and let δ be the diagonalization of α. So by definition δ is $\forall \mathsf{z}(\mathsf{Diag}(\overline{\ulcorner\alpha\urcorner},\mathsf{z}) \to \varphi(\mathsf{z}))$.

Because diagonalizing α yields δ, by definition $diag(\ulcorner\alpha\urcorner) = \ulcorner\delta\urcorner$. So since Diag expresses *diag*, we know that $\mathsf{Diag}(\overline{\ulcorner\alpha\urcorner},\overline{\ulcorner\delta\urcorner})$; indeed, $\mathsf{Diag}(\overline{\ulcorner\alpha\urcorner},\mathsf{z})$ is *only* satisfied by $\ulcorner\delta\urcorner$.

So $\forall \mathsf{z}(\mathsf{Diag}(\overline{\ulcorner\alpha\urcorner},\mathsf{z}) \to \varphi(\mathsf{z}))$ is true if and only if $\ulcorner\delta\urcorner$ satisfies $\varphi(\mathsf{z})$. In other words, δ is true if and only if $\varphi(\overline{\ulcorner\delta\urcorner})$ is true. □

Proof for (ii). Since by hypothesis Diag captures *diag* in T, (iv) from Defn. 37 just above tells us that *if* $diag(\ulcorner\alpha\urcorner) = \ulcorner\delta\urcorner$, *then* $T \vdash \forall \mathsf{z}(\mathsf{Diag}(\overline{\ulcorner\alpha\urcorner},\mathsf{z}) \leftrightarrow \mathsf{z} = \overline{\ulcorner\delta\urcorner})$.

But by definition $diag(\ulcorner\alpha\urcorner) = \ulcorner\delta\urcorner$. Hence $T \vdash \forall \mathsf{z}(\mathsf{Diag}(\overline{\ulcorner\alpha\urcorner},\mathsf{z}) \leftrightarrow \mathsf{z} = \overline{\ulcorner\delta\urcorner})$.

Since T can prove the equivalence of $\mathsf{Diag}(\overline{\ulcorner\alpha\urcorner},\mathsf{z})$ and $\mathsf{z} = \overline{\ulcorner\delta\urcorner}$, T can also prove the equivalence of $\forall \mathsf{z}(\mathsf{Diag}(\overline{\ulcorner\alpha\urcorner},\mathsf{z}) \to \varphi(\mathsf{z}))$ and $\forall \mathsf{z}(\mathsf{z} = \overline{\ulcorner\delta\urcorner} \to \varphi(\mathsf{z}))$.

In other words, $T \vdash \delta \leftrightarrow \forall \mathsf{z}(\mathsf{z} = \overline{\ulcorner\delta\urcorner} \to \varphi(\mathsf{z})))$. Which trivially gives us $T \vdash \delta \leftrightarrow \varphi(\overline{\ulcorner\delta\urcorner})$. □

A bit of jargon: by a mild abuse of mathematical terminology, we say

Defn. 52. *If* δ *is such that* $T \vdash \delta \leftrightarrow \varphi(\overline{\ulcorner\delta\urcorner})$*, then it is said to be a* fixed point *for* φ.

So the Diagonalization Lemma – or rather, part (ii) of it – is often called the Fixed Point Theorem: for appropriate theories T, every one-place open sentence has a fixed point.

13.2 Incompleteness from the Diagonalization Lemma

Suppose as usual that $Prf_T(m,n)$ is the relation which holds just if m numbers a T proof of a sentence with g.n. n (we continue to assume, of course, that we have a normal Gödel-numbering scheme in place). If T is p.r. axiomatized and contains Q, this relation is p.r. decidable and can be captured in T by a Σ_1 wff $\mathsf{Prf}_T(\mathsf{x},\mathsf{y})$. And now we pick up again an idea we first met in §3.6:

Defn. 16. *Put* $\mathsf{Prov}_T(\mathsf{y}) =_{\text{def}} \exists\mathsf{x}\mathsf{Prf}_T(\mathsf{x},\mathsf{y})$ *(where the quantifier runs over all the numbers in the domain). Then* $\mathsf{Prov}_T(\overline{\mathsf{n}})$ *says that some number Gödel-numbers a* T*-proof of the wff with Gödel-number* n*, i.e. the wff with code number* n *is a* T*-theorem. So* $\mathsf{Prov}_T(\mathsf{x})$ *is naturally called a* provability predicate.

Hence $\mathsf{Prov}_T(\overline{\ulcorner\varphi\urcorner})$ is true just when φ is a theorem. Note, by the way, that since Prov_T is built by existentially quantifying a Σ_1 wff, it is also Σ_1.

And here now is a general result about fixed points for the *negation* of such a provability predicate:

Theorem 48. *Suppose* T *is p.r. axiomatized, contains* Q*, and some sentence* γ *is a fixed point for* $\neg\mathsf{Prov}_T$*; in other words, suppose* $T \vdash \gamma \leftrightarrow \neg\mathsf{Prov}_T(\overline{\ulcorner\gamma\urcorner})$*. Then (i) if* T *is consistent,* $T \nvdash \gamma$*. And (ii) if* T *is* ω*-consistent,* $T \nvdash \neg\gamma$*.*

Again, to avoid unsightly rashes of subscripts, let's drop subscript 'T's. Then we can argue like this:

Proof. (i) Suppose $T \vdash \gamma$. Then, since $T \vdash \gamma \leftrightarrow \neg\mathsf{Prov}_T(\overline{\ulcorner\gamma\urcorner})$, we have $T \vdash \neg\mathsf{Prov}(\overline{\ulcorner\gamma\urcorner})$. But if there *is* a proof of γ, then for some m, $Prf(m, \ulcorner\gamma\urcorner)$, so $T \vdash \mathsf{Prf}(\overline{\mathsf{m}}, \overline{\ulcorner\gamma\urcorner})$, since T captures Prf by Prf. Hence $T \vdash \exists\mathsf{x}\,\mathsf{Prf}(\mathsf{x}, \overline{\ulcorner\gamma\urcorner})$, i.e. we also have $T \vdash \mathsf{Prov}(\overline{\ulcorner\gamma\urcorner})$, making T inconsistent. So if T is consistent, $T \nvdash \gamma$.

(ii) Suppose $T \vdash \neg\gamma$. Then, since $T \vdash \gamma \leftrightarrow \neg\mathsf{Prov}_T(\overline{\ulcorner\gamma\urcorner})$, we have $T \vdash \mathsf{Prov}(\overline{\ulcorner\gamma\urcorner})$, i.e. $T \vdash \exists\mathsf{x}\,\mathsf{Prf}(\mathsf{x}, \overline{\ulcorner\gamma\urcorner})$. Given T is consistent and proves $\neg\gamma$, there is no proof of γ, i.e. for every m, not-$Prf(m, \ulcorner\gamma\urcorner)$, whence for every m, $T \vdash \neg\mathsf{Prf}(\overline{\mathsf{m}}, \overline{\ulcorner\gamma\urcorner})$. So we have a $\psi(\mathsf{x})$ such that T proves $\exists\mathsf{x}\psi(\mathsf{x})$ while it refutes each instance $\psi(\overline{\mathsf{m}})$. Hence, if it is consistent, T is ω-inconsistent. So if T is ω-consistent (and hence consistent), $T \nvdash \neg\gamma$. □

But of course the general Diagonalization Lemma tells us that, as a special case,

Theorem 49. *If* T *is p.r. axiomatized, contains* Q*, then there exists a sentence* γ *such that* $T \vdash \gamma \leftrightarrow \neg\mathsf{Prov}_T(\overline{\ulcorner\gamma\urcorner})$*.*

Moreover, since Prov_T is Σ_1, $\neg\mathsf{Prov}_T$ is Π_1, and the diagonalization construction produces a Π_1 fixed point γ (see if you can work out why).

So putting those last two theorems together, we immediately recover Theorem 44.[1]

13.3 Proving our old G_T is in fact a fixed point for $\neg\mathsf{Prov}_T$

How does our new proof of the syntactic incompleteness theorem relate to the old one?

[1] Warning. Some authors call any fixed point γ for $\neg\mathsf{Prov}_T$ a Gödel sentence for T. That's fine, as long as you are alert to the fact that not everything that is true of Gödel sentences in the narrow sense we introduced in the preceding chapters is true of Gödel sentences in this wider sense. There's more about this in *IGT2*.

Our canonical Gödel sentence G_T, recall, was so constructed that it is true if and only if unprovable-in T. This fact can now be *expressed* inside T itself, by the wff $\mathsf{G}_T \leftrightarrow \neg\mathsf{Prov}_T(\overline{\ulcorner \mathsf{G}_T \urcorner})$. But T doesn't just express this fact but can quite easily *prove* it too, i.e. we have

Theorem 50. *If T is p.r. axiomatized and contains Q, $T \vdash \mathsf{G}_T \leftrightarrow \neg\mathsf{Prov}_T(\overline{\ulcorner \mathsf{G}_T \urcorner})$.*

In other words, G_T is one of those fixed points for $\neg\mathsf{Prov}_T$ which are all undecidable by T. And note that (notation apart) this gives us the key result we simply announced way back in §3.6. The proof is a bit fiddly but elementary. Dropping subscripts once more:

Proof. By definition, $Gdl(m, n)$ holds iff $Prf(m, diag(n))$. We can therefore fix on the following canonical definition:

$$\mathsf{Gdl}(\mathsf{x}, \mathsf{y}) =_{\text{def}} \exists \mathsf{z}(\mathsf{Prf}(\mathsf{x}, \mathsf{z}) \wedge \mathsf{Diag}(\mathsf{y}, \mathsf{z})).$$

And now let's do some elementary manipulations:

$$\begin{array}{lll}
\mathsf{G} =_{\text{def}} \forall \mathsf{x} \neg \mathsf{Gdl}(\mathsf{x}, \overline{\ulcorner \mathsf{U} \urcorner}) & & \\
\quad \leftrightarrow \quad \forall \mathsf{x} \neg \exists \mathsf{z}(\mathsf{Prf}(\mathsf{x}, \mathsf{z}) \wedge \mathsf{Diag}(\overline{\ulcorner \mathsf{U} \urcorner}, \mathsf{z})) & & \text{(definition of Gdl)} \\
\quad \leftrightarrow \quad \forall \mathsf{x} \forall \mathsf{z} \neg (\mathsf{Prf}(\mathsf{x}, \mathsf{z}) \wedge \mathsf{Diag}(\overline{\ulcorner \mathsf{U} \urcorner}, \mathsf{z})) & & \text{(pushing in the negation)} \\
\quad \leftrightarrow \quad \forall \mathsf{z} \forall \mathsf{x} \neg (\mathsf{Prf}(\mathsf{x}, \mathsf{z}) \wedge \mathsf{Diag}(\overline{\ulcorner \mathsf{U} \urcorner}, \mathsf{z})) & & \text{(swapping quantifiers)} \\
\quad \leftrightarrow \quad \forall \mathsf{z}(\mathsf{Diag}(\overline{\ulcorner \mathsf{U} \urcorner}, \mathsf{z}) \rightarrow \neg \exists \mathsf{x}\, \mathsf{Prf}(\mathsf{x}, \mathsf{z})) & & \text{(rearranging after '}\forall \mathsf{z}\text{')} \\
\quad =_{\text{def}} \forall \mathsf{z}(\mathsf{Diag}(\overline{\ulcorner \mathsf{U} \urcorner}, \mathsf{z}) \rightarrow \neg \mathsf{Prov}(\mathsf{z})) & & \text{(definition of Prov)}
\end{array}$$

Since this is proved by simple logical manipulations, that means we can prove the equivalence inside the formal first-order logic built into Q and hence in T. Therefore we have

$$T \vdash \mathsf{G} \leftrightarrow \forall \mathsf{z}(\mathsf{Diag}(\overline{\ulcorner \mathsf{U} \urcorner}, \mathsf{z}) \rightarrow \neg \mathsf{Prov}(\mathsf{z})).$$

Now, diagonalizing U yields G. Hence, just by the definition of *diag*, we have $diag(\ulcorner \mathsf{U} \urcorner) = \ulcorner \mathsf{G} \urcorner$. Since by hypothesis Diag captures *diag*, it follows that

$$T \vdash \forall \mathsf{z}(\mathsf{Diag}(\overline{\ulcorner \mathsf{U} \urcorner}, \mathsf{z}) \leftrightarrow \mathsf{z} = \overline{\ulcorner \mathsf{G} \urcorner}).$$

Putting those two results together, we immediately get

$$T \vdash \mathsf{G} \leftrightarrow \forall \mathsf{z}(\mathsf{z} = \overline{\ulcorner \mathsf{G} \urcorner} \rightarrow \neg \mathsf{Prov}(\mathsf{z})).$$

But the right-hand side of that biconditional is trivially equivalent to $\neg\mathsf{Prov}(\overline{\ulcorner \mathsf{G} \urcorner})$. So we've proved the desired result. □

We should note a simple corollary of that last theorem. Suppose Prov_T not only expresses but *captures* the property of being a T-theorem. Then, by the definition of capturing, if φ is a non-theorem, then we'd have $T \vdash \neg\mathsf{Prov}_T(\overline{\ulcorner \varphi \urcorner})$. And in particular, since G_T is a non-theorem, $T \vdash \neg\mathsf{Prov}_T(\overline{\ulcorner \mathsf{G}_T \urcorner})$. But, given Theorem 50, that would imply $T \vdash \mathsf{G}_T$ which we know is false.

Hence Prov_T does *not* capture the property of being a T-theorem.

13.4 Rosser's Theorem

One half of the syntactic First Theorem requires the assumption that we are dealing with a theory T which is not only consistent but is ω-consistent. But we can improve on this. Following Barkley Rosser, we can construct a Rosser sentence R_T such that we only need to assume T is plain consistent in order to show that R_T is formally undecidable.

So how does Rosser do the trick? Essentially, where Gödel constructs a sentence G_T that indirectly says 'I am unprovable in T', Rosser constructs R_T which indirectly says 'if I am provable in T, then my negation is already provable' (i.e. it says that if there is a proof of R_T with super g.n. n, then there is a proof of $\neg\mathsf{R}_T$ with a smaller code number).

How do we formally implement this idea? Start by considering the relation $\overline{Prf}_T(m, n)$ which holds when m numbers a T-proof of the *negation* of the wff with number n. This relation is obviously p.r. given that Prf_T is; so assuming T has the usual properties it will be captured by a wff $\overline{\mathsf{Prf}}_T(\mathsf{x}, \mathsf{y})$. So we can construct *the Rosser provability predicate* defined as follows:

Defn. 53. $\mathsf{RProv}_T(\mathsf{x}) =_{\text{def}} \exists\mathsf{v}(\mathsf{Prf}_T(\mathsf{v}, \mathsf{x}) \wedge (\forall\mathsf{w} \leq \mathsf{v})\neg\overline{\mathsf{Prf}}_T(\mathsf{w}, \mathsf{x}))$.

Then a sentence is Rosser-provable in T – its g.n. satisfies the Rosser provability predicate – if it has a proof (in the ordinary sense) and there's no 'smaller' proof of its negation.

Now we apply the Diagonalization Lemma, not to the negation of a regular provability predicate (which is what we just did to get Gödel's First Theorem again), but to the negation of the Rosser provability predicate. The Lemma then tells us,

Theorem 51. *Given that T is p.r. axiomatized and contains Q, then there is a sentence R_T such that $T \vdash \mathsf{R}_T \leftrightarrow \neg\mathsf{RProv}_T(\overline{\ulcorner\mathsf{R}_T\urcorner})$.*

We call such a sentence R_T a Rosser sentence for T. We can now show that

Theorem 52. *Let T be consistent p.r. axiomatized theory which contains Q, and let ρ be any fixed point for $\neg\mathsf{RProv}_T(\mathsf{x})$. Then $T \nvdash \rho$ and $T \nvdash \neg\rho$.*

And since the Diagonalization Lemma tells us that there is such a fixed point, it follows that T has an undecidable sentence R_T, without now requiring ω-consistency. Moreover, with extra effort, we can get ourselves a Π_1 sentence which is undecidable if T is consistent. Which gives us Rosser's Theorem in the form:

Theorem 53. *Let T be a consistent p.r. axiomatized theory which contains Q, then there is Π_1 sentence φ such that $T \nvdash \varphi$ and $T \nvdash \neg\varphi$.*

But what we gain on the swings (a slightly stronger result) we lose on the roundabouts (in the complications in getting there). There's no avoiding it; the

proof of Theorem 52 is decidedly unpretty. We then have to do more work to beef up that proof idea to show that in fact (as with Gödel's original proof) we can find a Π_1 sentence which is undecidable so long as T is consistent.[2] But you do not need to know these proofs! – it is just good to know that they exist.

Note, though, that once more Rosser's Theorem gives us an incompletability result. Suppose T is a consistent p.r. axiomatized theory which contains Q. It is incomplete. Add more axioms. It will remain incomplete unless it becomes inconsistent or stops being p.r. axiomatized.

[2]Masochists can check out the proof of Theorem 25.4 and further details in *IGT2*.