11 The First Incompleteness Theorem, semantic version

Let's quickly review some crucial ideas that we've met in the last couple of chapters:

i. We fixed on a particular scheme for coding up wffs of PA's language L_A by using Gödel numbers ('g.n.' for short), and for coding up PA-proofs by super Gödel numbers (assuming for convenience that these proofs are simple sequences of wffs). It is crucial that the algorithms which take us from expressions to code numbers and back again don't involve any open-ended searches. Call a coding scheme with this feature *normal*. (§10.2)

ii. Notation: If φ is an expression, then we'll denote its Gödel number in our logician's English by '$\ulcorner\varphi\urcorner$'. We use '$\overline{\ulcorner\varphi\urcorner}$' as an abbreviation inside L_A for the standard numeral for $\ulcorner\varphi\urcorner$. (§10.4)

iii. $Prf(m, n)$ is the relation which holds just if m is the super g.n. (on our normal scheme) of a sequence of wffs that is a PA proof of a sentence with g.n. n. Crucially, this relation is primitive recursive. (§10.2)

iv. Any p.r. function or relation can be *expressed* by a wff of PA's language L_A. In particular, we can choose a Σ_1 wff which 'canonically' expresses *Prf* by recapitulating its p.r. definition (or more strictly, by recapitulating the definitional chain for the relation's characteristic function). (§9.3)

v. Any p.r. function or relation can be *captured* in Q and hence in PA. In particular, *Prf* can be captured by a Σ_1 wff (and again, one which recapitulates the relation's p.r. definition). (§9.5)

For what follows, it isn't necessary that you know the *proofs* of the claims we've just summarized: but do pause to check that you at least fully understand what the various claims *mean*.

In this chapter, we now learn how to construct a 'Gödel sentence' for PA more or less as announced in §3.6. Then we use this in a proof that PA is incomplete, given the semantic assumption that the theory is sound. We then show how to generalize the result to other theories. In the following chapter, we use our Gödel

sentence to prove that PA is incomplete, but this time on the basis of a syntactic assumption. Again, the argument will generalize.

11.1 The idea of diagonalization

As we announced right back in Chapter 3, Gödel is going to tell us how to construct a wff G in PA that is true if and only if it is unprovable in PA. We now have an inkling of how he can do that: wffs can contain numerals which refer to numbers which – via Gödel coding – are in turn correlated with wffs. Maybe, if we are cunning, we can get a wff to be – via the coding – 'about' itself. And the next section shows how.

But first we need to introduce a simple but pivotal construction. We will say

Defn. 44. *The diagonalization of a wff φ with one free variable is the wff* $\varphi(\overline{\ulcorner\varphi\urcorner})$.

That is to say, making the free variable explicit, the diagonalization of the wff $\varphi(\mathsf{x})$ is what you get by substituting for its free variable x the numeral for the Gödel number of the wff $\varphi(\mathsf{x})$.

Why is this substitution operation called *diagonalization*? Well compare the 'diagonal' construction we encountered in §4.3. There, we counted off wffs $\varphi_0(\mathsf{x})$, $\varphi_1(\mathsf{x})$, $\varphi_2(\mathsf{x})\ldots$ in an enumeration of wffs with one free variable; and then we substituted (the numeral for) the index n for the free variable in the wff φ_n, to form $\varphi_n(\overline{\mathsf{n}})$. We can now think of the Gödel number of a wff as indexing that wff in a list of wffs. And so, in our new diagonal construction, we are again substituting (the numeral for) the index of a wff for the free variable in the wff.

Diagonalization is a simple mechanical operation on expressions. Hence,

Theorem 32. *There is a p.r. function $diag(n)$ which, when applied to a number n which is the g.n. of some L_A wff with one free variable, yields the g.n. of that wff's diagonalization, and yields n otherwise.*

Proof. Consider this procedure. Try treating n as a g.n., and seek to decode it. If you don't get an expression with one free variable, return n. Otherwise you get a wff of the type $\varphi(\mathsf{x})$, and can then form the wff $\varphi(\overline{\mathsf{n}})$, which is its diagonalization (since by assumption n is its g.n.). Then we work out the g.n. of this resulting wff to compute $diag(n)$.

This procedure doesn't involve any unbounded searches. So we again will be able to program the procedure using just 'for' loops. Hence *diag* is a p.r. function. □

And, for future reference, since Q expresses and captures every p.r. function, that means in particular we have

Theorem 33. *There is wff* $\mathsf{Diag}(\mathsf{x},\mathsf{y})$ *which (canonically) expresses and captures diag in* Q.

11.2 Constructing a Gödel sentence

Recall, the relation $Prf(m, n)$ is defined to hold just when m is the super g.n. for a PA proof of the wff with g.n. n. Now for a simple tweak:

Defn. 45. *The relation $Gdl(m, n)$ is defined to hold just when m is the super g.n. for a* PA *proof of the diagonalization of the wff with g.n. n.*

Theorem 34. *Gdl is primitive recursive.*

Official proof. $Gdl(m, n)$ holds when $Prf(m, diag(n))$. Let c_{Prf} be the characteristic function of Prf, which is primitive recursive. Then $c_{Gld}(x, y) =_{\text{def}} c_{Prf}(x, diag(y))$ will be the characteristic function of Gdl, so – being a composition of p.r. functions – c_{Gdl} is primitive recursive too. □

Informal proof. We just remark that, as with $Prf(m, n)$, we can mechanically check whether $Gdl(m, n)$. Just decode m. Check whether it gives a sequence of wffs. If it does, check whether it is a PA proof. If it is, ask whether the concluding wff of the proof is the result of taking an open wff whose g.n. is n and diagonalizing that wff. That involves no open-ended searches. An algorithm involving only 'for' loops will suffice. So Gdl is primitive recursive. □

Since Gdl is p.r. it can be both expressed and captured in PA by a canonical Σ_1 wff, i.e. one which recapitulates a definition of the (characteristic function) of the p.r. relation. So let's say

Defn. 46. $\mathsf{Gdl}(\mathsf{x}, \mathsf{y})$ *stands in for some Σ_1 wff which canonically expresses and captures Gdl.*

And now comes the simple but ingenious Gödelian construction! First we form the open wff we'll abbreviate as U, or to make its free variable explicit,

Defn. 47. $\mathsf{U}(\mathsf{y}) =_{\text{def}} \forall \mathsf{x} \neg \mathsf{Gdl}(\mathsf{x}, \mathsf{y})$.

Then we diagonalize U, to give

Defn. 48. $\mathsf{G} =_{\text{def}} \mathsf{U}(\overline{\ulcorner \mathsf{U} \urcorner}) = \forall \mathsf{x} \neg \mathsf{Gdl}(\mathsf{x}, \overline{\ulcorner \mathsf{U} \urcorner})$.

And here is the wonderful result:

Theorem 35. G *is true if and only if it is unprovable in* PA.

Proof. Consider what it takes for G to be true (on the interpretation built into L_A of course), given that the formal predicate Gdl expresses the numerical relation Gdl.

G is true if and only if for all numbers m it isn't the case that $Gdl(m, \ulcorner \mathsf{U} \urcorner)$. That is to say, given the definition of Gdl, G is true if and only if there is no number m such that m is the code number for a PA proof of the diagonalization

of the wff with g.n. $\ulcorner \mathsf{U} \urcorner$. But the wff with g.n. $\ulcorner \mathsf{U} \urcorner$ is of course U; and its diagonalization is G.

So, G is true if and only if there is no number m such that m is the code number for a PA proof of G. But if G is provable, some number would indeed be the code number of a proof of it. Hence G is true if and only if it is unprovable in PA. □

Pause and admire this ingeniously elegant construction!

G – meaning of course the L_A sentence you get when you unpack the abbreviations! – is thus our promised Gödel sentence for PA. We might indeed call it a *canonical* Gödel sentence for three reasons: (a) it is defined in terms of a wff that we said canonically expresses/captures *Gdl*, and (b) because it is roughly the sort of sentence that Gödel himself constructed, so (c) it is the kind of sentence people standardly have in mind when they talk of '*the*' Gödel sentence for PA.

It is true that G will be horribly long when spelt out in quite unabbreviated L_A. But in another way, it is relatively simple. We have the easy result that

Theorem 36. G *is* Π_1.

Proof. $\mathsf{Gdl}(\mathsf{x}, \mathsf{y})$ is Σ_1. So $\mathsf{Gdl}(\mathsf{x}, \overline{\ulcorner \mathsf{U} \urcorner})$ is still Σ_1 (we've just filled up one slot in the open wff with a numeral). So its negation $\neg\mathsf{Gdl}(\mathsf{x}, \overline{\ulcorner \mathsf{U} \urcorner})$ is Π_1 (since the negation of a Σ_1 wff is Π_1). Hence $\forall \mathsf{x} \neg \mathsf{Gdl}(\mathsf{x}, \overline{\ulcorner \mathsf{U} \urcorner})$ is Π_1 too (since the result of adding another universal quantifier on the front of a Π_1 wff is still Π_1). □

11.3 What G says

It is often claimed that a Gödel sentence like G actually *says* of itself that it is unprovable. However, that can't be strictly true.

G (when unpacked) is just another sentence of PA's language L_A, the language of basic arithmetic. It is a long wff involving the first-order quantifiers, the connectives, the identity symbol, and 'S', '$+$' and '$\times$', which all have the standard interpretation built into L_A. In particular, the standard numeral in G refers to a number, not a wff; and the quantifier in G runs over numbers. So G strictly speaking says something about *numbers*, not about wffs and their unprovability.

However there is perhaps a reasonable sense in which G *can* be described as *indirectly* saying that it is unprovable. Note, this is *not* to make play with some radical re-interpretation of G's symbols (for doing *that* would just make any claim about what G says boringly trivial: if we are allowed radical re-interpretations – like spies choosing to borrow ordinary words for use in a secret code – then any string of symbols can be made to say anything). No, it is because the symbols are still being given their *standard* interpretation that we can recognize that Gdl (when unpacked) will express *Gdl*, given the background framework of Gödel numbering which is involved in the definition of the relation *Gdl*. Therefore, given that coding scheme, we can recognize just from its construction that G will be true when no number m is such that $Gdl(m, \overline{\ulcorner \mathsf{U} \urcorner})$, and so no number

numbers a proof of G. In short, given the coding scheme, we can see just from the way it is constructed that G that it is true just when it is unprovable. *That* is the limited sense in which, via our Gödel coding, the canonical G signifies or indirectly says that it is unprovable.

11.4 The First Theorem for PA – the semantic version

We already announced in §3.6 that Gödel tells us how to construct a wff which is true if and only if unprovable.[1] And as we showed then, the argument to an incompleteness theorem is now very straightforward. Here it is again.

Assume PA is a sound theory, i.e. it proves no falsehoods (because its axioms are true and its logic is truth-preserving). If G could be proved in PA, then PA *would* prove a false theorem (since G is true if and only if it is *not* provable). That would contradict our supposition that PA is sound. Hence, G is not provable in PA.

But that shows that G *is* true. So ¬G must be false. Hence ¬G cannot be proved in PA either, supposing PA is sound.

In Gödel's words, then, G is a 'formally undecidable' sentence of PA (see Defn. 7):

Theorem 37. *If* PA *is sound, then there is a true* Π_1 *sentence* G *such that* $\mathsf{PA} \nvdash \mathsf{G}$ *and* $\mathsf{PA} \nvdash \neg\mathsf{G}$*, so* PA *is negation incomplete.*

If we are happy with the semantic assumption that PA's axioms *are* true on interpretation and so PA *is* sound, the argument for incompleteness is as simple as that – or at least, it's that simple once we have constructed G.

11.5 Generalizing the proof

This line of proof now generalizes.

Suppose T is any theory which (i) contains the language of basic arithmetic (see Defn. 11), so T can form standard numerals, and we can form the diagonalization of a T-wff with one free variable. Suppose (ii) that we working with a normal system of Gödel-numbering for T-wffs and T-proofs. Suppose also (iii) that T is p.r. axiomatized in the sense of Defn. 42. Then we can again define a relation $Gld_T(m, n)$ which holds when m numbers (on our new scheme) a T-proof of the diagonalization of the wff with number n; and this relation will be primitive recursive again.

Continuing to suppose that T's language includes the language of basic arithmetic, T will be able to express the p.r. relation Gld_T by a Σ_1 wff Gld_T. Then, just as we did for PA, we'll be able to construct the corresponding Π_1 wff G_T. And by exactly the same argument as before we can show, more generally,

[1] In fact, if you have been paying close attention to details, you'll spot that we haven't yet *quite* joined up the treatment here with that earlier sketch. But we get all the way there in §13.3.

Theorem 38. *If T is a sound p.r. axiomatized theory whose language contains the language of basic arithmetic, then there will be a true Π_1 sentence G_T such that $T \nvdash \mathsf{G}_T$ and $T \nvdash \neg\mathsf{G}_T$, so T is negation incomplete.*

Which is at last our first, '*semantic*', version of the generalized First Incompleteness Theorem!

Let's note an immediate corollary:

Theorem 39. *There is no p.r. axiomatized theory framed in the language of L_A whose theorems are all and only the truths of L_A.*

For if the theorems are all true, our theory is sound, and hence it can't be complete.

11.6 Our Incompleteness Theorem is better called an *incompletability* theorem

Here, we just repeat the argument of §2.3: but the point is crucial enough to bear repetition. Suppose T is a sound p.r. axiomatized theory which can express claims of basic arithmetic. Then by Theorem 38 we can find a true G_T such that $T \nvdash \mathsf{G}_T$ and $T \nvdash \neg\mathsf{G}_T$. That *doesn't* mean that G_T is 'absolutely unprovable' in any sense: it just means that G_T-is-unprovable-in-T.

OK: let's just augment T by adding G_T as a new axiom, to give the theory $U = T + \mathsf{G}_T$. Then (i) U is still sound (for the old T-axioms are true, the added new axiom is true, and the logic is still truth-preserving). (ii) U is evidently still a p.r. axiomatized theory (why?). (iii) We haven't changed the language. So our Incompleteness Theorem applies, and we can find a sentence G_U such that $U \nvdash \mathsf{G}_U$ and $U \nvdash \neg\mathsf{G}_U$. Since U is can prove everything T can prove, that implies $T \nvdash \mathsf{G}_U$ and $T \nvdash \neg\mathsf{G}_U$. In other words, as we put it before, 'repairing the gap' in T by adding G_T as a new axiom leaves some other sentences that are undecidable in T *still* undecidable in the augmented theory.

In sum, our incompleteness theorem tells us that if we keep chucking more and more additional axioms at T, our theory will still remain negation-incomplete, unless it either stops being sound or stops being p.r. axiomatized. In a good sense, T is *incompletable*.

11.7 Comparing old and new semantic incompleteness theorems

Compare the new Theorem 38 which we *have* proved with the old theorem which we initially announced in §2.2 but of course *didn't* there prove:

Theorem 1. *Suppose T is a formal axiomatized theory whose language contains the language of basic arithmetic. Then, if T is sound, there will be a true 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 is stronger in one respect, weaker in another. But the gain is much more than the loss.

Our new theorem is stronger, because it tells us much more about the character of the undecidable Gödel sentence – namely it has minimal quantifier complexity. The unprovable sentence G_T is a Π_1 sentence of arithmetic. We'll return to say more about this in the next chapter.

Our new theorem is weaker, however, as it only applies to p.r. axiomatized theories, not to (effectively) axiomatized theories more generally. But that's no real loss. Indeed as we will see in §12.5, Gödel's own original theorems only strictly speaking apply to p.r. axiomatized theories. And after all, what would a theory look like that was effectively axiomatized but *not* p.r. axiomatized? It would mean that we could e.g. only tell what's an axiom on the basis of an open-ended search: but that would require an *entirely* unnatural way of specifying the theorem's axioms in the first place. As we noted at the end of §10.3, any normally presented effectively axiomatized theory will be p.r. axiomatized.

So while we *can* beef up our new result to make it apply as generally as the announced Theorem 1 – and we will say more about this in the next Interlude – there is really only limited interest in doing so. The real force of the (semantic) First Incompleteness Theorem is already captured by Theorem 38.