2 The First Theorem, two versions

2.1 Soundness, consistency, etc.

Let's read into the record two standard definitions:

Defn. 9. *A theory T is* sound *iff its axioms are true (on the interpretation built into T's language), and its logic is truth-preserving, so all its theorems are true.*

Defn. 10. *A theory T is (syntactically)* consistent *iff there is no φ such that $T \vdash \varphi$ and $T \vdash \neg\varphi$, where '$\neg$' is T's negation operator.*

In a classical setting, if T is inconsistent, then $T \vdash \psi$ for all ψ. And of course, soundness implies consistency. We shouldn't need to delay over these no doubt familiar ideas.

But we also need another natural enough definition to use in this chapter:

Defn. 11. *The formalized interpreted language L* contains the language of basic arithmetic *if L has a term which denotes zero and function symbols for the successor, addition and multiplication functions defined over numbers – these can be either built-in as primitives or introduced by definition – and has the usual connectives, the identity predicate, and can express quantifiers running over the natural numbers.*

An example might be the language of set theory, in which we can define zero, successor, addition and multiplication in standard ways, and express restricted quantifiers running over just zero and its successors.

(OK, you might worry whether the natural number system referred to in set theory is the genuine article or just a structurally equivalent surrogate. But then what is 'the genuine article'? We are not going tangle with *that* messy issue, as we have quite enough other things to worry about! When we talk, then, of a theory quantifying over numbers, then, take it to be quantifying over numbers or to whatever surrogates can play the role of natural numbers. Nothing relevant to our project hangs on the difference.)

2.2 Two theorems distinguished

In his 1931 paper, Gödel proves (or rather more accurately, gives us most of the materials to prove) the following:

Theorem 1. *If T is a sound formal axiomatized theory whose language contains the language of basic arithmetic, then 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.*

We will outline a pivotal part of Gödel's proof (in a very gappy way!) in the next chapter.

However this version of an incompleteness theorem *isn't* what is most commonly referred to as *the* First Theorem, nor is it the result that Gödel foregrounds in his 1931 paper. For note, Theorem 1 tells us what follows from a *semantic* assumption, namely the assumption that T is sound. And soundness is defined in terms of truth.

Now, post-Tarski, most of us aren't particularly scared of the notion of the truth. To be sure, there are issues about how best to treat the notion formally, to preserve as many as possible of our pre-formal intuitions while e.g. blocking the Liar Paradox. But most of us think that we don't have to regard the relevant notion of a sound theory as metaphysically loaded in an obscure and worrying way. But Gödel was writing at a time when, for various reasons (think logical positivism!), the very idea of truth-in-mathematics was under some suspicion. So it was *extremely* important to Gödel that he could show that you don't need to deploy any semantic notions to get an incompleteness result. So he demonstrates the following:

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

Being consistent (in the relevant sense) is a syntactic property; being able to formally prove enough arithmetic is another syntactic property; and the mysterious additional property which I haven't explained yet is synactically defined too. So *this* version of the incompleteness theorem only makes syntactic assumptions.

Of course, we'll need to be a lot more explicit in due course; but this indicates the general *character* of Gödel's central result. And remembering the title of his 1931 paper, 'can prove a modest amount of arithmetic' is what it takes for a theory to be sufficiently related to *Principia*'s for the theorem to apply. But I'll not pause to spell out just how much arithmetic that is, though we'll eventually find that it is stunningly little. Nor will I pause to explain that 'additional property' condition. We'll meet it in due course, but also explain how – by a cunning trick discovered by J. Barkley Rosser in 1936 – we can drop that condition again.

For now, then, the first important take-away message of this chapter is that the incompleteness theorem does come in two flavours. There's a version making a semantic assumption (an assumption about what the relevant theory T can *express*), and there's a version making a syntactic assumption (an assumption about what T can *prove*). It is important to keep this firmly in mind.

2.3 Incompleteness and incompletability

Let's concentrate on the first, semantic, version of the First Theorem.

Suppose T is a sound theory which contains the language of basic arithmetic. Then, the claim is, we can find a true G_T such that $T \nvdash \mathsf{G}_T$ and $T \nvdash \neg\mathsf{G}_T$. Let's be really clear: this doesn't, repeat *doesn't*, say that G_T is 'absolutely unprovable', whatever that could possibly mean. It just says that that G_T and its negation are unprovable-in-T.

Ok, you might well ask, why don't we just 'repair the gap' in T by adding the true sentence G_T as a new axiom? Well, consider the theory $U = T + \mathsf{G}_T$ (to use an obvious notation). Then (i) U is still sound, since the old T-axioms are true, the added new axiom is true, and its logic is still truth-preserving. (ii) U is still a properly formalized theory, since adding a single specified axiom to T doesn't make it undecidable what is an axiom of the augmented theory. (iii) U still contains the language of basic arithmetic. So Theorem 1 still applies, and we can find a sentence G_U such that $U \nvdash \mathsf{G}_U$ and $U \nvdash \neg\mathsf{G}_U$. And since U is stronger than T we have, a fortiori, $T \nvdash \mathsf{G}_U$ and $T \nvdash \neg\mathsf{G}_U$. In other words, '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.

And so it goes. Keep throwing more and more additional true axioms at T and our theory still remains negation-incomplete, unless it stops being effectively axiomatized. So here's the second important take-away message of the chapter: when the conditions for Theorem 1 apply, then the theory T will not just be incomplete but in a good sense T will be *incompletable*. (We'll see in due course that just the same holds when the conditions Theorem 2 apply.)

So we should perhaps really talk of the First *Incompletability* Theorem.

2.4 The completeness and incompleteness theorems

A reality check. We've already made the distinction we need in §1.3, and we illustrated it then with a toy example. But experience suggests that it will do no harm at all to hammer home the point!

A *semantic completeness theorem* of the kind you are no doubt familiar with from elementary logic courses is about the relation between semantic and syntactic consequence relations. In particular, you will know about this result:

> If T is a theory cast in a first-order language with a standard first-order deductive apparatus, then for any φ, if $T \vDash \varphi$ then $T \vdash \varphi$.

The deductive logic is complete because every conclusion that semantically follows from some given premisses can be deduced from them by formal logical proof. Compare now our current semantic version of the *incompleteness theorem*, applied in particular to a theory with a first-order logic:

> If T is a sound theory with a standard first-order logic that can express enough arithmetic, then T is not negation complete, i.e. there

is some φ such that neither $T \vdash \varphi$ nor $T \vdash \neg\varphi$.

Compare the two statements carefully.

If T is a sound arithmetical theory with a standard first-order logic (for example) then it will have a deductive system which is semantically complete in the sense of tracking all the genuine first-order logical consequences. Yet obviously, such a theory can easily fail to be negation-complete in the same way as our toy example in §1.3, e.g. by quite boringly missing out some absolutely basic arithmetical axiom!

The First Incompleteness Theorem tells us that, much less boringly, try as we might, *every* theory of arithmetic satisfying certain desirable conditions (even if it has a semantically complete logic) will remain negation incomplete as a theory.