

Gödel Without Tears – 1

Incompleteness – the very idea

Peter Smith

October 16, 2009

Why these notes? After all, I've written *An Introduction to Gödel's Theorems* (CUP, heavily corrected fourth printing 2009: henceforth *IGT*). Surely that's more than enough to be going on with?

Ah, but there's the snag. It *is* more than enough. In the writing, as is the way with these things, the book grew far beyond the scope of the lecture notes from which it started. And while I hope the result is still pretty accessible to someone prepared to put in the time and effort, there's a *lot* more in the book than is really needed by philosophers meeting the incompleteness theorems for the first time. After all, you might want to get your heads around only those technical basics which are actually needed for understanding philosophical discussions about incompleteness.

So you need a cut-down version of the book – an introduction to the *Introduction!* Well, isn't that what lectures are for? Indeed. But there's another snag. I haven't got many lectures to play with. So either (A) I crack on at quite a fast pace (hard-core mathmo style), cover those basics, but perhaps leave too many people puzzled and alarmed. Or (B) I do relaxed talk'n'chalk, highlighting the really Big Ideas, making sure everyone is grasping them as we go along, but inevitably omit important stuff and leave quite a gap between what happens in the lectures and what happens in the book. What to do?

I'm going for plan (B). But then I still need to do something to fill that gap between lectures and book. Hence these notes. The idea, then, is to give relaxed lectures, highlighting Big Ideas, not worrying too much about depth or fine-detail (or even about getting through *all* of the day's intended menu of topics). Then after the lecture, I'll write up notes that expand things just enough, and then give pointers to relevant chunks of *IGT*. The idea, however, is for the notes to be more or less stand-alone, and to tell a brief but coherent story read by themselves. So occasionally I'll copy a paragraph or two from the book, rather than just refer to them.

Warning: just occasionally in these notes, I'll no doubt apply that good maxim 'Where it doesn't itch, don't scratch'. In other words, sometimes I'll say things that are not utterly rigorous, but I hope in unworrying ways. If you are bright enough to spot the slight cheats or corner-cutting, you should be bright enough to spot how to repair what I say, at the cost of a bit of fuss and bother, so no harm done!

1 Kurt Gödel (1906–1978)

The greatest logician of the twentieth century. Born in what is now Brno. Educated in Vienna. At 23, his doctoral dissertation established the completeness of the first-order predicate calculus. Later he would do immensely important work on set theory, as well as make contributions to proof theory. Talk of 'Gödel's Theorems', however, typically refers to his two incompleteness theorems in an epoch-making 1931 paper. Left Austria for the USA in 1938, and spent rest of his life at the Institute of Advanced Studies at Princeton. Always a perfectionist, after the mid 1940s he more or less stopped publishing.

For a brief overview of his life and work, see http://en.wikipedia.org/wiki/Kurt_Gödel, or better – though you'll need to skip – <http://plato.stanford.edu/entries/goedel>. There's a nice biography, John Dawson *Logical Dilemmas* (A. K. Peters, 1997), which will also give you some sense of the logical scene in the glory days of the 1930s.

2 ‘On formally undecidable propositions of *Principia Mathematica* and related systems I’

This is the title of the 1931 paper which proves the First Incompleteness Theorem and states the Second Theorem. (The ‘I’ indicates that it is the first part of what was going to be a two part paper, with Part II spelling out the proof of the Second Theorem. But that was never written. I’ll explain later why Gödel didn’t need to bother.)

Even the title gives us a number of things to explain. What’s a ‘formally undecidable proposition’? What’s *Principia Mathematica*? – you’ve heard of it, no doubt, but what’s the project of that triple-decker work? What counts as a ‘related system’? In fact, what is meant by ‘system’ here? Take the last question first.

2.1 ‘Systems’ – i.e. formal axiomatized theories

Our concern is with formal axiomatized theories. T is such a theory if it has (i) a formalized interpreted language L , (ii) an effectively decidable set of axioms, (iii) a formalized proof-system in which we can deduce theorems from the axioms.

To explain, we first need a definition:

Defn. 1. A property P defined over a domain D is effectively decidable iff there’s an algorithm for settling, for any $o \in D$, whether o has property P – i.e. there’s a step-by-step mechanical routine for settling the issue, a suitably programmed computer could do the trick. A set Σ is effectively decidable if the property of being a member of that set is effectively decidable.

(i) We’ll take the idea of a formal interpreted language L to be familiar from earlier logic courses. There will be a *syntax* which fixes which strings of symbols form terms, which form wffs, and in particular which strings of symbols form sentences, i.e. closed wffs with no unbound variables dangling free. There will be a *semantics* which gives the intended interpretation of L , fixing truth conditions for each L -sentence. Crucially, to emphasize,

Defn. 2. For a formalized language L , the syntactic rules of L must be such that the properties of being a term, a wff, a wff with one free variable, a sentence, etc. are effectively decidable.

For what’s the point of having a formal syntax if we can’t then decide whether a string of symbols is or is not a wff, for example?

(ii) A theory T built in language L will have a certain class of L -sentences picked out as *axioms*. Again it is to be effectively decidable what’s an axiom. (After all, if we build a theory and then can’t routinely tell whether a given sentence is one of its axioms, what would be the point of that?)

(iii) Just laying down a bunch of axioms would normally be pretty idle if we can’t deduce conclusions from them! So a formal axiomatized theory T comes equipped with a proof-system, a set of rules for deducing further theorems from our initial axioms. Again, describing a proof-system such that we couldn’t routinely tell whether its rules are being followed would have little point. Hence we naturally require that it is effectively decidable whether a given array of wffs is indeed a proof according to the rules. But it doesn’t matter whether the proof-system is a Frege/Hilbert axiomatic logic, a natural deduction system, a tree/tableau system – so long as it is indeed effectively checkable that a candidate proof-array has the property of being properly constructed according to the rules.

So, in summary

Defn. 3. A formal axiomatized theory T has a formalized language L , a certain class of L -sentences picked out as axioms where it is decidable what’s an axiom, and it has a proof-system such that it is effectively decidable whether a given array of wffs is indeed a proof according to the rules.

Careful!!! To say that, for a properly formalized theory T it must be effectively decidable whether a purported T -proof of φ is indeed a kosher proof according to T ’s deduction system is not, repeat *not*, to say that it must be effectively decidable whether φ has a proof. It is one thing

to be able to effectively *check* a proof once found, it is another thing to be able to effectively *decide in advance* whether there exists a proof to be discovered.

Finally, a couple of useful notational conventions should be mentioned for future reference.

1. Greek letters, as in the ‘ φ ’ we’ve just used, are schematic variables in the metalanguage in which we talk about our formal systems.
2. Particular expressions from formal systems – and abbreviations of them – will be in sans serif type.

For more explanations, see *IGT*, §§2.2, 3.1–3.3, 4.1.

2.2 ‘Formally undecidable propositions’ and negation incompleteness

Defn. 4. ‘ $T \vdash \varphi$ ’ says: there is a formal deduction in T ’s proof-system from T -axioms to the sentence φ as conclusion. If φ is a sentence and $T \vdash \varphi$, then φ is said to be a theorem of T .

So NB, ‘ \vdash ’ officially signifies provability in T , a formal syntactically definable relation, not semantic entailment.

Defn. 5. If T is a formal theory, and φ is some sentence of the language of that theory, then T formally decides φ iff either $T \vdash \varphi$ or $T \vdash \neg\varphi$.

Hence,

Defn. 6. A sentence φ is formally undecidable by T iff $T \not\vdash \varphi$ and $T \not\vdash \neg\varphi$.

Another bit of terminology:

Defn. 7. A theory T is negation complete iff it formally decides every closed wff of its language – i.e. for every sentence φ , $T \vdash \varphi$ or $T \vdash \neg\varphi$.

Trivially, then, there are ‘formally undecidable propositions’ in T if and only if T isn’t negation complete.

Of course, it is very easy to construct negation-incomplete theories: just leave out some seemingly necessary basic assumptions about the matter in hand! But suppose we are trying to fully pin down some body of truths using a formal theory. We fix on an interpreted formal language L apt for expressing such truths. And then we’d ideally like to build a theory T in L , whose axioms are such that when (but only when) φ is true, $T \vdash \varphi$. So, making the classical assumption that either φ is true or $\neg\varphi$ is true, we’d like T to be such that either $T \vdash \varphi$ or $T \vdash \neg\varphi$. Negation completeness, then, is a natural desideratum for theories.

For more explanations, see *IGT*, §3.4.

2.3 Deductivism, logicism, and *Principia*

The basic arithmetic of successor (‘next number’), addition, and multiplication is child’s play (literally!). It is entirely plausible to suppose that, whether the answers are readily available to us or not, questions posed in the language of basic arithmetic – the language of successor, addition, and multiplication plus familiar first-order logical apparatus – have entirely determinate answers. These answers are surely ‘fixed’ by (a) the basic zero-and-its-successors structure of the natural number series and (b) the nature of addition and multiplication as given by the school-room explanations.

So it is surely plausible to suppose that we should be able lay down a bunch of axioms which characterize the number series, addition and multiplication (which codify what we teach the kids), and that these axioms should settle every truth of basic arithmetic, in the sense that every such truth of the language of successor, addition, and multiplication is logically provable from the axioms. For want of a standard label, call this view *deductivism* about basic arithmetic.

What could be the status of the axioms? You might, for example, be a Kantian deductivist who holds that the axioms encapsulate ‘intuitions’ in which we grasp the fundamental structure of

the numbers and the nature of addition and multiplication, where these ‘intuitions’ are a special cognitive achievement in which we somehow represent to ourselves the arithmetical world.

But talk of intuition can be very puzzling and problematic. So we might prefer Frege’s view that the axioms are *analytic*, truths of logic or rather of logic-plus-definitions. On this view, we don’t need Kantian ‘intuitions’ going beyond logic. The Fregean brand of deductivism is standardly dubbed ‘logicism’.

Famously, Frege’s attempt to be a logicist deductivist about arithmetic hit the rocks, because Russell showed that his logical system is inconsistent in a pretty elementary way (it is beset by Russell’s Paradox). That devastated Frege, but Russell was undaunted, and still gripped by deductivist ambitions he wrote:

All mathematics [yep! – *all* mathematics] deals exclusively with concepts definable in terms of a very small number of logical concepts, and . . . all its propositions are deducible from a very small number of fundamental logical principles.

That’s a big promisory note in Russell’s *The Principles of Mathematics* (1903). His attempt (with Whitehead) to make good on that promise is *Principia Mathematica* (three volumes, though unfinished, 1910, 1912, 1913). The project is to set down some logical axioms and definitions and deduce the laws of arithmetic from them. Famously, they eventually get to prove that $1 + 1 = 2$ at *110.643 (Volume II, page 86), accompanied by the wry comment, ‘The above proposition is occasionally useful’.

2.4 Gödel’s bomb

There are technical complications which means that *Principia*’s axioms are not all clearly ‘logical’ (in particular there’s an appeal to a brute-force *Axiom of Infinity* which in effect states that there is an infinite number of objects, and to a very dodgy *Axiom of Reducibility*: for more on this see <http://plato.stanford.edu/entries/principia-mathematica/>). But leave those worries aside – they pale into insignificance compared with the bomb exploded by Gödel. For his First Incompleteness Theorem shows that any form of deductivism about basic arithmetic (not just *Principia*’s) is in trouble.

Why? Well the proponent of deductivism about basic arithmetic (logicist or otherwise) wants to pin down first-order arithmetical truths about successor/addition/multiplication, without leaving any out: so he wants to give a negation-complete theory. *And there can’t be such a theory.* Gödel’s First Theorem says – at a very rough first shot – that *nice theories containing enough arithmetic are always negation incomplete.*

So varieties of deductivism, and logicism in particular, must always fail – a major (and perhaps very surprising) result!

‘Hold on! I’ve heard of neo-logicism which has its enthusiastic advocates. How can that be so if Gödel showed that logicism is a dead duck?’ Well, we might still like the idea that some logical principles plus what are more-or-less definitions together *semantically* entail all arithmetical truths, while allowing that we can’t capture the relevant entailment relation in a single properly axiomatized deductive system of logic. Then the resulting overall system of arithmetic won’t count as a formalized theory of all arithmetical truth since its logic is not formalizable, and Gödel’s theorems don’t apply. But more about that in due course.

2.5 The First Incompleteness Theorem just a bit more carefully, in two versions

Three more definitions:

Defn. 8. *The formalized language L contains the language of basic arithmetic if L has at least the standard first-order logical apparatus, has a term ‘0’ which denotes zero and function symbols for the successor, addition and multiplication functions defined over numbers – either built-in as primitives or introduced by definition – and has a predicate whose extension is the natural numbers.*

The point of that last clause is that if ‘N’ is a predicate satisfied just by numbers, then the wff $\forall x(Nx \rightarrow \varphi(x))$ says that every number satisfies φ ; so L can make general claims specifically about natural numbers. (If L is already defined to be a language whose quantifiers run over the numbers, then you can just use ‘ $x = x$ ’ for ‘N’.)

Defn. 9. A theory T is sound if its axioms are true (on the interpretation built in to T ’s language), and its logic is truth-preserving, so all its theorems are true.

Defn. 10. A theory T is consistent if there is no φ such that $T \vdash \varphi$ and $T \vdash \neg\varphi$.

In a classical setting, if T is inconsistent, then $T \vdash \psi$ for all ψ . And of course, trivially, soundness implies consistency.

Gödel now proves (more accurately, gives us most of the materials to prove) the following:

Theorem 1. If T is a sound formalized theory whose language contains the language of basic arithmetic, then there will be a true sentence G_T of basic arithmetic such that $T \not\vdash G_T$ and $T \not\vdash \neg G_T$, so T must be negation incomplete.

However that *isn’t* what is usually referred to as the First Incompleteness Theorem. For note, Theorem 1 is about what follows from a *semantic* assumption, namely that T is sound. And soundness is defined in terms of truth. Now, post-Tarski, we aren’t scared of the notion of the truth! But Gödel was writing at a time when, for various reasons, the very idea of truth-in-mathematics was under some suspicion. So it was *extremely* important to him to show that you don’t need to deploy any semantic notions to get (again roughly) the following result:

Theorem 2. For any consistent formalized theory T which contains a certain modest amount of arithmetic (and has a certain additional desirable property that any sensible formalized arithmetic will share), there is a sentence of basic arithmetic G_T such that $T \not\vdash G_T$ and $T \not\vdash \neg G_T$, so T must be negation incomplete.

Of course, we’ll need to be a lot more explicit in due course, but that gives the flavour of the result. The ‘contains a modest amount of arithmetic’ is what makes a theory sufficiently related to *Principia*’s for the theorem to apply. I’ll not pause in this lecture to spell out that just how much arithmetic that is, but we’ll find that it is stunningly little. (Nor will I pause now to explain that ‘additional desirable property’ condition. We’ll meet it in due course, but also explain how – by a cunning trick discovered by J. Barkley Rosser in 1936 – how we can drop that condition.)

For the present, however, let’s concentrate on the semantic version of Gödel’s theorem, i.e. Theorem 1.

2.6 Theorem 1 is better called an *incompleteness* theorem

Suppose T is a sound theory which can express claims of basic arithmetic. Then we can find a true G_T such that $T \not\vdash G_T$ and $T \not\vdash \neg G_T$.

Of course, that *doesn’t* mean that G_T is ‘absolutely unprovable’, whatever that could mean. It just means that G_T -is-unprovable-in- T .

Now, we might want to ‘repair the gap’ in T by adding G_T as a new axiom. So consider the theory $U = T + G_T$ (to use an obvious notation). 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 still a properly formalized theory, since adding an specified axiom to T doesn’t make it undecidable what is an axiom of the augmented theory. (iii) U still can express claims of basic arithmetic. So Gödel’s First Incompleteness Theorem applies, and we can find a sentence G_U such that $U \not\vdash G_U$ and $U \not\vdash \neg G_U$. And since U is stronger than T , we have a fortiori, $T \not\vdash G_U$ and $T \not\vdash \neg G_U$. 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 chucking more and more additional true axioms at T and our theory still remains negation-incomplete, unless it stops being sound or stops being effectively axiomatizable. In a good sense, T is *incompletable*.

3 How did Gödel prove the First Theorem (in the semantic version)?

Let's take a first pass at outlining how Gödel proved the semantic version of his incompleteness theorem. Obviously we'll be coming back to this in a lot more detail later, but we can give just a flavour of what's going on. We kick off two natural definitions.

Defn. 11. *If L contains the language of basic arithmetic, so it contains a term 0 for zero and a function expression S for the successor function, then the terms $0, S0, SS0, SSS0, \dots$, are L 's standard numerals, and we'll use 'n' to abbreviate the standard numeral for n .*

Henceforth, we'll assume that the language of any theory we are interested in contains the language of basic arithmetic and hence has standard numerals denoting the numbers.

Defn. 12. *The formal wff $\varphi(x)$ of the interpreted language L expresses the numerical property P iff $\varphi(n)$ is true on interpretation just when n has property P . Similarly, the formal wff $\psi(x, y)$ expresses the numerical relation R iff $\psi(m, n)$ is true just when m has relation R to n .*

Then the proof in outline form goes as follows:

1. *Set up a Gödel numbering* We are nowadays familiar with the idea that all kinds of data can be coded up using numbers. So suppose we set up a sensible (effective) way of coding wffs and sequences of wffs by natural numbers – so-called Gödel-numbering. Then, given a formalized theory T , we can define e.g. the numerical properties Wff_T , $Sent_T$, Prf_T and $Prov_T$, where

$Wff_T(n)$ iff n is the code number of a T -wff.

$Sent_T(n)$ iff n is the code number of a T -sentence.

$Prf_T(m, n)$ iff m is the code number of a T -proof of the T -sentence with code number n .

$Prov_T(n)$ iff n is the code number of T -theorem.

2. *Expressing such properties/relations inside T* We next show that such properties/relations can be expressed inside T by wffs of the formal theory belonging to the language of basic arithmetic [takes a bit of work!]. We show in particular how to build an arithmetic wff we'll abbreviate $Prov_T(x)$ that expresses the property $Prov_T$, so $Prov_T(n)$ is true exactly when $Prov_T(n)$, i.e. when n is the code number of T -theorem.
3. *The construction: building a Gödel sentence* Next – the really cunning bit, but surprisingly easy – we show how to build a 'Gödel' sentence G_T such that G_T is in fact equivalent to $\neg Prov_T(g)$, where the numeral 'g' denotes the code-number for G_T . In other words, G_T is true if and only if G_T isn't a theorem.
4. *The argument* Suppose $T \vdash G_T$. Then G_T would be provable, and hence G_T would be false, so T would have a false theorem and hence not be sound, contrary to hypothesis. So $T \not\vdash G_T$. So G_T is true. So $\neg G_T$ is false and T , being sound, can't prove it. Hence we also have $T \not\vdash \neg G_T$.

There are big gaps to fill, but that's the strategy. (The proof of Theorem 2 then shows that we can get the same result using the same construction of a Gödel sentence by dropping the assumption that T is sound, so long as we require a bit more by way of what the theory T can prove, and require T to have that currently mysterious 'additional desirable property'. More about this in due course)

Of course, you might immediately think something a bit worrying about our sketch so far. For basically, I'm saying we can construct an arithmetic sentence in T that, via the Gödel number coding, says 'I am not provable in T '. But shouldn't we be suspicious about that? After all, we know we get into paradox if we try to play with sentences that say 'I am not true'. So why does the self-reference in the Liar sentence lead to *paradox*, while the self-reference in Gödel's proof give us a *theorem*? A very good question. I hope that over the coming lectures, the answer to that good question will become clear!

Now read *IGT*, §§1.1–3.4.