Gödel Without (Too Many) Tears – 1

Incompleteness – the very idea

Peter Smith

University of Canterbury, Christchurch, NZ

April 7, 2010

- The notion of effective decidability
- What's a formalized language?
- What's a formal axiomatized theory?
- What's negation incompleteness?
- 'Deductivism' about basic arithmetic
- Two versions of Gödel's First Incompleteness Theorem
- The First Theorem as an *incompletability* theorem
- How did Gödel prove (one version of) the First Incompleteness Theorem?

Why these notes? After all, I've already written a pretty detailed book, *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 is – to be frank – a *lot* more in the book than is really needed by philosophers meeting the incompleteness theorems for the first time, or indeed by mathematicians wanting a brisk introduction. You might reasonably want to get your heads around only those technical basics which are actually necessary for understanding how the theorems are proved and for appreciating 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 a very 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 ought 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 (nor even about getting through *all* of the day's intended menu of topics). These notes then expand things just enough, and give pointers to relevant chunks of *IGT*. Though I hope these notes will be to a fair extent be stand-alone, and 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. And the notes come with a Logical Health Warning: in the interests of relative brevity, I'll occasionally have to 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 that can be easily remedied if you are feeling pernickety.

The bullet-pointed headers to each helping of notes – to each episode, as I'll call it – give pointers/reminders to the coverage.

A final introductory remark. If you notice any typos/thinkos in these notes and/or the latest printing of the first edition of *IGT* please let me know (peter_smith@me.com). In due course, there will be a second edition of the book; so I'd also be very grateful for any more general comments about the book that might help me improve the book. Some further relevant materials, plus the latest version of these notes, can be found at www.logicmatters.net.

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* theorem for the first-order predicate calculus (i.e. a standard proof system for first-order logic indeed captures all the valid inferences – where validity is defined semantically). Later he would do immensely important work on set theory, as well as make contributions to proof theory. He even later wrote on models of General Relativity. Talk of 'Gödel's Theorems', however, typically refers to his two *incompleteness* theorems in an epoch-making 1931 paper.[1]

Gödel 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 very nice biography, John Dawson *Logical Dilemmas* (A. K. Peters, 1997), which will also give you a real 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, just what is meant by 'system' here? We'll take the last question first.

2.1 'Systems' – i.e. formal axiomatized theories

Our concern is with systems in the sense of *formal axiomatized theories*. T is such a theory if it has (i) an effectively formalized language L, (ii) an effectively decidable set of axioms, (iii) an effectively 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 in a finite number of steps, 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 in principle do the trick. A set Σ is effectively decidable if the property of being a member of that set is effectively decidable.*

Now take in turn those conditions (i) to (iii) for being a formal axiomatized theory.

(i) We'll assume that the general idea of a formal L is familiar from earlier logic courses. There will be a *syntax* which fixes which strings of symbols form terms, which form wffs, and

[1] Yes, Gödel proved a 'completeness theorem' and 'incompleteness theorems'. By the end of this first episode you should be able to tell the difference!

in particular which strings of symbols form *sentences*, i.e. closed wffs with no unbound variables dangling free. And crucially, to emphasize what is perhaps not emphasized in introductory courses,

Defn. 2. *For an effectively formalized language L, the basic alphabet of L is to be finite, and 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.*

NB, the restriction to a finite basic vocabulary still allows us, e.g., to have an infinite supply of variables: for example, given the two symbols 'x' and '$'$' we can construct an infinite supply of composite symbols $\mathsf{x}, \mathsf{x}', \mathsf{x}'', \mathsf{x}'''$.[2] As to the effective decidability of the properties of being a term, etc., the point of setting up a formal language is usually (inter alia) precisely to put issues of what is and isn't a sentence beyond dispute, so we want to be able effectively to decide whether a string of symbols is or is not a sentence. A formal interpreted language will also normally have an intended *semantics* which gives the interpretation of L, fixing truth conditions for each L-sentence – again, the semantics should be presented in such a way that we can mechanically read off from the interpretation rules the interpretation of any given sentence.

(ii) A theory T built in language L will have a certain class of L-sentences picked out as *axioms*.[3] Again it is to be *effectively decidable* what's an axiom. (After all, if we are making a theory rigorous, but then can't routinely tell whether a given sentence is one of its axioms, that would – usually – be pretty pointless.)

(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. But describing a proof-system such that we couldn't then routinely tell whether its rules are in fact being followed wouldn't have much point. Hence we naturally require that it is effectively decidable whether a given array of wffs is indeed a proof from the axioms according to the rules. It doesn't matter for our purposes whether the proof-system is e.g. 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 an effectively 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 from the axioms according to the rules.*

Careful, though! To say that, for a properly formalized theory T it must be effectively decidable whether a given 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 proposed, it is another thing to be able to effectively *decide in advance* whether there is exists a proof to be discovered. (It will turn out, for example, that any formal axiomatized theory T containing a certain modicum of arithmetic is such that, although you can mechanically check a purported proof of φ to see whether it *is* a proof, there's no general way of telling of an arbitrary φ whether it is provable in T or not.)

2.2 Notational conventions

Before going on, we should highlight a couple of useful notational conventions that we'll be using from now on in these notes (the same convention is used in *IGT*, and indeed is not an uncommon one):

[2] In some contexts, for technical purposes, there is interest in talking about formal languages with an uncountably infinite number of primitives. That's why the finiteness constraint needs to be made explicit.

[3] We'll allow the class of axioms to be null. It should be familiar that we can trade in axioms for rules of inference – though we can't trade in all rules of inference for axioms if we want to be able to deduce anything: cf. Lewis Carroll's Achilles and the Tortoise!

1. Particular expressions from formal systems – and abbreviations of them – will be in sans serif type. Examples: $\mathsf{SS0} + \mathsf{S0} = \mathsf{SSS0}$, $\forall \mathsf{x}\, \mathsf{Sx} \neq \mathsf{0}$. [Blackboard convention: overline formal wffs when clarity demands. Bracketing will tend to be casual.]

2. Expressions in informal mathematics will be in ordinary serif font (with variables, function letters etc. in italics). Examples: $2 + 1 = 3$, $n + m = m + n$, $S(x + y) = x + Sy$.

3. Greek letters, as in the 'φ' we've just used, are schematic variables in the metalanguage in which we talk about our formal systems.

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

2.3 '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 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 $T \nvdash \varphi$ and $T \nvdash \neg\varphi$.*

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 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.

2.4 Deductivism, logicism, and *Principia*

The elementary 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 what we'll call *the language of basic arithmetic* – i.e. 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 fundamental zero-and-its-successors structure of the natural number series (with zero not being a successor, every number having a successor, distinct numbers having distinct successors, and so the sequence of zero and its successors never circling round but marching off for ever) plus (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 these axioms. For want of a standard label, call this view *deductivism* about basic arithmetic.

What could be the status of the axioms? I suppose 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 is very puzzling and problematic. So we might well be tempted instead by 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: logical reasoning alone is enough. The Fregean brand of deductivism is standardly dubbed 'logicism'.

Famously, Frege's attempt to be a logicist deductivist about arithmetic (in fact, for him, more than basic arithmetic) hit the rocks, because – as Russell showed – his logical system is in fact 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). And *Principia Mathematica* (three volumes, though unfinished, 1910, 1912, 1913) is Russell's attempt with Whitehead to make good on that promise. The project is to set down some logical axioms and definitions and deduce the laws of basic arithmetic (and then more) 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'.

2.5 Gödel's bomb

Principia, frankly, is a bit of a mess – in terms of clarity and rigour, it's quite a step backwards from Frege. And there are technical complications which mean that not all *Principia*'s axioms are clearly 'logical' even in a stretched sense. 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 then there is the notoriously dodgy *Axiom of Reducibility*.[4]. But leave those worries aside – they pale into insignificance compared with the bomb exploded by Gödel.

For Gödel's First Incompleteness Theorem shows that any form of deductivism about even just 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 basic arithmetic are always negation incomplete.*

So varieties of deductivism, and logicism in particular, must always fail. Which is a rather stunning result![5]

[4] *Principia* without the dodgy Axiom is a 'type theory' which is quite nicely motivated, but you can't reconstruct much maths in it: the dodgy Axiom of Reducibility allows you to reconstruct classical maths by pretending that the type distinctions by which we are supposed to avoid paradox can be ignored when we need to do so: for more on this see http://plato.stanford.edu/entries/principia-mathematica/

[5] '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 formal axiomatized theory of all arithmetical truth since its logic is not formalizable, and Gödel's theorems don't apply.

3 The First Incompleteness Theorem, a bit more carefully

3.1 Two versions of the First Theorem

Three more definitions. First, let's be a bit more careful about that idea of 'the language of basic arithmetic':

Defn. 8. *The formalized language L* contains the language of basic arithmetic *if L has at least the standard first-order logical apparatus (including identity), has a term '$\mathsf{0}$' 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 $\forall \mathsf{x}(\mathsf{Nx} \to \varphi(\mathsf{x}))$ 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 could use '$\mathsf{x} = \mathsf{x}$' for 'N', or – equivalently – just forget about it!)

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 $T \vdash \neg\varphi$,*

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, 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 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.*

However that *isn't* what is usually referred to as the First Incompleteness Theorem. For note, Theorem 1 tells us 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 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 blocking versions of the Liar Paradox. But most of us think that we don't have to regard the general idea of truth 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. There were other reasons too for wanting to steers away from semantic notions, reasons to do with 'Hilbert's program' about which more anon. So it was *extremely* important to Gödel to show that you don't need to deploy any semantic notions to get (again roughly) the following result:

Theorem 2. *For any consistent formal axiomatized 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 $T \nvdash \mathsf{G}_T$ and $T \nvdash \neg\mathsf{G}_T$, so T is negation incomplete.*

(Here 'contains' means not just can express but *can prove.*) Of course, we'll need to be a lot more explicit in due course, but that indicates the general character of Gödel's result. The 'contains a modest amount of arithmetic' is what makes a theory sufficiently related to *Principia*'s for the theorem to apply – remember the title of Gödel's paper! I'll not pause in this first episode 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.

3.2 Theorem 1 is better called an *incompletability* theorem

Suppose T is a sound theory which can express claims of basic arithmetic. Then we can find a true G_T such that $T \nvdash \mathsf{G}_T$ and $T \nvdash \neg\mathsf{G}_T$. *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 $U = T + \mathsf{G}_T$ (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 $U \nvdash \mathsf{G}_U$ 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 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*.

4 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 with 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* $\mathsf{0}$, $\mathsf{S0}$, $\mathsf{SS0}$, $\mathsf{SSS0}$, *…, are L's* standard numerals*, and we'll use '*$\overline{\mathsf{n}}$*' to abbreviate the standard numeral for n.*

(The overlining convention to indicate standard numerals is a pretty standard one.) 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* $\varphi(\mathsf{x})$ *of the interpreted language L* expresses *the numerical property P iff* $\varphi(\overline{\mathsf{n}})$ *is true on interpretation just when n has property P. Similarly, the formal wff* $\psi(\mathsf{x},\mathsf{y})$ *expresses the numerical relation R iff* $\psi(\overline{\mathsf{m}},\overline{\mathsf{n}})$ *is true just when m has relation R to n. And the formal wff* $\chi(\mathsf{x},\mathsf{y})$ *expresses the numerical function f iff* $\chi(\overline{\mathsf{m}},\overline{\mathsf{n}})$ *is true just when* $f(m) = n$.

The generalization to many-place relations/many-argument functions is obvious.

Then the proof of Theorem 1 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 formal axiomatized theory T, we can define e.g. the numerical properties Wff_T, Sent_T, Prf_T and Prov_T, where

   $\mathit{Wff}_T(n)$ iff n is the code number of a T-wff.
   $\mathit{Sent}_T(n)$ iff n is the code number of a T-sentence.
   $\mathit{Prf}_T(m,n)$ iff m is the code number of a T-proof of the T-sentence with code number n.
   $\mathit{Prov}_T(n)$ 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 – just out of the materials of the language of basic arithmetic – an arithmetic formal wff we'll abbreviate $\mathsf{Prov}_T(\mathsf{x})$ that expresses the property Prov_T, so $\mathsf{Prov}_T(\overline{\mathsf{n}})$ is true exactly when $\mathit{Prov}_T(n)$, 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 $\neg\mathsf{Prov}_T(\overline{\mathsf{g}})$, where the standard numeral '$\overline{\mathsf{g}}$' is the numeral denoting the code-number for G_T. In other words (think about it!!), G_T is true if and only if G_T isn't a theorem.

4. *The argument* Suppose $T \vdash \mathsf{G}_T$. Then G_T would be a theorem, and hence G_T would be false, so T would have a false theorem and hence not be sound, contrary to hypothesis. So $T \nvdash \mathsf{G}_T$. So G_T is true. So $\neg\mathsf{G}_T$ is false and T, being sound, can't prove it. Hence we also have $T \nvdash \neg\mathsf{G}_T$.

There are big gaps to fill there, but that's the overall 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. 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 episodes, the answer to that good question will become clear!

Now read *IGT*, §§1.1–3.4. ("But hold on! What's the Second Theorem that you mentioned?" Good question – the opening chapter of *IGT* will tell you: but here in these notes we'll maintain the suspense as far as Episode 10!)