Gödel's Theorem: A Proof from The Book?

Peter Smith

March 10, 2006

1 Intro

Here's one version Gödel's 1931 First Incompleteness Theorem:

> If T is a nice, sound theory of arithmetic, then it is incomplete, i.e. there are arithmetical sentences φ such that T proves neither φ nor $\neg\varphi$.

There are three things here to explain straight away:

1. A theory is *sound* if all its theorems are true (typically, this is because the theory has true axioms and truth-preserving inference rules).

2. We don't mean anything exotic here by a *theory of arithmetic*. We just mean that T can at least talk about natural numbers and can express generalizations about them. Also T can at least talk about the successor function which takes us from one number to the next, and can talk about addition, multiplication, and exponentiation. (Maybe T can talk about a great deal more.)

3. As to *nice*, that's just casual shorthand here for the assumption that T is a properly constructed formal theory. That is to say, in particular, that it is mechanically decidable what is to count as an axiom of the theory. And it is mechanically decidable what is to count as a well-constructed formal derivation of some theorem from the axioms according to T's proof-system.

Gödel's Theorem is an astonishing result. It means you can't regiment arithmetical truth into a nice sound theory T. There will always be an arithmetical φ sentence such that T proves neither φ nor $\neg\varphi$. But one of those sentences is true, hence there's an arithmetical truth such that T can't prove it.

So, just for a start, bang goes the Frege/Russell logicist programme of trying to show that all true arithmetic (indeed, all classical analysis) can be derived from a few logical principles plus some crafty definitions. Such a programme can't ever be completed: any nice sound theory we construct will miss out on some truths.

So how do we prove Gödel's Theorem? Seventy-something years on, we know a variety of significantly different ways of doing the job. But what are the nicest, most elegant, most insightful ways? The great mathematician Paul Erdős had the fantasy of a Book in which God keeps the best proofs of mathematical theorems. What proofs of Gödel's result belong in The Book? A good question! This talk sketches one Book proof.

2 Nice theories

Let's start with a couple of easy results about nice, properly axiomatized theories.[1] First

Theorem 1 *The theorems of a properly axiomatized theory T are effectively enumerable.*

Proof A set is 'effectively enumerable' if it is either empty or there is a computer program that (in principle, when run for ever!) lists its members in some order, repetitions allowed. Here's how to write a program to spit out all the theorems of T. Write a program that generates in turn – in some kind of 'alphabetical order' – the strings of symbols of T's alphabet. By hypothesis, if T is properly axiomatized, we'll be able to mechanically select out those strings that are properly formed proofs in T's logic, which only appeal to axioms in T's list of axioms. So that means we can mechanically select out those strings which are proofs in the theory T. Whenever we find such a proof, print out the proof's conclusion. And that will give us a mechanically generated list of T's theorems. ⊠

And here's a converse result:

Theorem 2 *For any effectively enumerable set of sentences Σ, there is a properly axiomatized theory T whose theorems are the members of Σ.*

Proof For the null case where Σ is empty take T to be the empty theory with no axioms and no rules of inference!

Otherwise, suppose the members of Σ can be effectively listed off $\varphi_1, \varphi_2, \varphi_3, \varphi_4, \ldots$. Now consider the following theory T.

1. Its axioms are $\varphi_1, \varphi_2 \wedge \varphi_2, \varphi_3 \wedge \varphi_3 \wedge \varphi_3, \varphi_4 \wedge \varphi_4 \wedge \varphi_4 \wedge \varphi_4, \ldots$.

2. Its sole proof-building rule is that a deduction is always a two-line affair, consisting of an axiom followed by one of the conjuncts in the axiom (there is a single inference move eliminating the conjunctions): and this two-line structure counts as a proof of its second line.

OK, that's a bit strange, but T is a perfectly well defined axiomatic theory. For (1) it is effectively decidable what is an axiom. Just inspect a wff: see if it is an n-fold conjunction $\psi \wedge \psi \wedge \ldots \psi$. If it isn't, it isn't an axiom. If it is, count n, then enumerate the φ_i, and see whether $\psi = \varphi_n$.[2]

And (2) it is effectively decidable what's a proof. Just check it indeed has the form of an axiom followed by one of its conjuncts.

But T's theorems, by construction, are exactly the φ_i. So we are done. ⊠

[1] In the talk as delivered, I had an aside about second-order Peano Arithmetic before this section, explaining why Gödel's Theorem didn't apply to it. But this proved unnecessarily distracting. So I've deleted that section here.

[2] Why didn't we just take the axioms of T to be the φ_i? Because we couldn't then mechanically decide whether a given wff ψ is an axiom. True, if ψ is one of the φ_i we could show it to be by listing off $\varphi_1, \varphi_2, \varphi_3, \varphi_4, \ldots$ and noting that ψ eventually turns up. But if ψ isn't an axiom, then this strategy of running through the φ_i wouldn't demonstrate the fact.

3 Gödel's Theorem and the truths of arithmetic

Given our first two theorems we quickly get another result:

Theorem 3 *Gödel's Incompleteness Theorem is equivalent to the claim that the truths of arithmetic can't be effectively enumerated.*

Proof (i) Suppose the truths of arithmetic *could* be effectively enumerated (that's the truths couched in the language of successor, addition, and multiplication, plus some standard logical apparatus). Then by Theorem 2, there would be a properly axiomatized theory T that proved all and only these truths. So given an arithmetic sentence φ, T would prove either φ or $\neg\varphi$, whichever is the truth. So T would be a sound, complete theory.

Hence, contraposing, if Gödel's Theorem is right, i.e. if there can't be a complete properly axiomatized and sound theory T, the truths can't be effectively enumerated.

(ii) Now for the other direction. Suppose the truths of arithmetic *can't* be effectively enumerated. But we know from Theorem 1 that the theorems of a given formalized arithmetic T *can* be effectively enumerated. So there is a mismatch between the truths and the T-theorems.

Assuming we are dealing with a sound formalized arithmetic T, T proves no falsehoods. The mismatch between the truths and the T-provable sentences must therefore be due to the there being truths which T can't prove. Suppose φ is one of these. Then T doesn't prove φ; and since $\neg\varphi$ is false, T doesn't prove that either. So T is incomplete. ⊠

So Theorem 3 signposts one route to establishing Gödel's Theorem: namely, prove that the truths of arithmetic can't be effectively enumerated. That's the route we are going to take.

It is worth saying that this wasn't Gödel's route in 1931, by the way. It couldn't have been because the general theory of computation, and so the general theory of effective enumerability, wasn't in place until about 1936.

4 Some more background

So to get Gödel's Theorem by our route, we want to establish that the truths of arithmetic can't be effectively enumerated.

Well, we can't get something for nothing, so we'll need *some* background stuff to work with. But we can get by with surprisingly little.

First, a notational preliminary. An arithmetic theory will be able to frame a series of expressions '$\mathsf{0}, \mathsf{S0}, \mathsf{SS0}, \mathsf{SSS0}, \ldots$'. Call these the standard numerals. We'll abbreviate the numeral for n by $\overline{\mathsf{n}}$.

Next, a remark on coding tricks. We can, in particular, use a numerical coding scheme to assign numerical codes to code up computer programs. A nice way of doing it is to use powers of primes. We associate symbols in the program with numbers; then we code up a string of n symbols by taking the first n primes, and raising each prime to the number of the associated symbol. The fact that a formal arithmetic knows about the exponential function means it can easily handle facts about codings. (Indeed that's exactly why I'm focusing on arithmetics with the exponential function available, though we can strictly speaking do without using a trick found by Gödel.)

Now let's state three easy but important facts:

Fact 1. Suppose R is a decidable relation among numbers (i.e. suppose that there is a mechanical way of deciding whether Rmn for given any numbers m, n). Then we can construct a corresponding formal expression φ using just addition, multiplication and exponentiation such that Rmn if and only if the formal sentence $\varphi(\overline{\mathsf{m}}, \overline{\mathsf{n}})$ is true. In other words, using limited resources we can construct a φ which is a formal expression for R.

Fact 2. If Σ is effectively enumerable, then there is some decidable relation R such that $n \in \Sigma$ if and only if $\exists s Rsn$.

Fact 3. We can effectively enumerate (the recipes for generating) the effectively enumerable sets of numbers. So we can effectively list off those sets $W_0, W_1, W_2, \ldots$.

Why do these facts hold?

Proof sketch for Fact 1 If R is decidable, that means there is a computer program which tells us whether Rmn (use some all-purpose programming language). So Rmn if and only if the decision program with inputs m, n terminates with a 'yes' verdict. But now let's use coding tricks to encode this sort of fact about our computer program into corresponding arithmetical statements. That way, we'll get an arithmetical statement involving $\overline{\mathsf{m}}, \overline{\mathsf{n}}$ which holds just when Rmn.

Proof sketch for Fact 2 Suppose some mechanical procedure P effectively enumerates Σ. Step through this procedure one step at a time; and consider the relation Rsn that holds just in case at step s the procedure P spits out the number n. Then R is decidable (there is a mechanical method of deciding whether P spits out n at step s – just run procedure P for s steps and see what happens!). And trivially, P eventually spits out n if and only if for some step-number s, Rsn holds.

Proof sketch for Fact 3 The third of our facts holds because we can effectively enumerate possible computer programs (in some all-purpose programming language) for spitting out lists of numbers.

5 Proving the non-effective enumerability of arithmetical truths

So, with those Facts under our belt, here again is the pivotal result we want to establish:

Theorem 4 *The truths of arithmetic can't be effectively enumerated.*

Proof Recall Fact 3, and consider an effective enumeration of the effectively enumerable numerical sets $W_0, W_1, W_2, \ldots$.

Now define the set $K =_{\text{def}} \{e \mid e \in W_e\}$ and its complement $\overline{K} =_{\text{def}} \{e \mid e \notin W_e\}$. Then here are two facts about this pair of sets:

1. K is effectively enumerable. You enumerate it by a kind of zigzag computation. Start be examining the first two members of W_0, then the first of W_1, then the third member of W_0, the second of W_1, the first of W_2, then the next members of W_0, W_1, W_2, and the first of W_3, and so on. At each step as you go along, when you look at another member of W_e, you check whether it is e, and if it is, you print it out. Eventually, every e such that $e \in W_e$ will be printed out.

2. $\overline{K}$ is not effectively enumerable. For any k, $k \in \overline{K}$ if and only if $k \notin W_k$. Hence, $\overline{K}$ cannot be identical to any of the W_k, so $\overline{K}$ isn't one of the effectively enumerable sets.

Now, because K is effectively enumerable, that means by Fact 2 there is a some decidable relation R such that $n \in K$ if and only if $\exists s R(s, n)$. Now, since R is decidable there is by Fact 1 a formal two-place expression R of the language of arithmetic which expresses R. In particular, then, we'll have

$n \in K$ if and only if $\exists \mathsf{s} \mathsf{R}(\mathsf{s}, \overline{\mathsf{n}})$ is true;
$n \in \overline{K}$ if and only if $\neg\exists \mathsf{s} \mathsf{R}(\mathsf{s}, \overline{\mathsf{n}})$ is true.

Concentrate on the second equivalence. If the truths of arithmetic $\mathscr{T}$ were effectively enumerable, we'd be able run through $\mathscr{T}$, and whenever we came across sentence of the type $\neg\exists \mathsf{s} \mathsf{R}(\mathsf{s}, \overline{\mathsf{n}})$, we could put n on a list of members of $\overline{K}$, and that way get an effective enumeration of $\overline{K}$. But we seen that $\overline{K}$ is *not* effectively enumerable. So $\mathscr{T}$ isn't effectively enumerable. Which is what we wanted to show. ⊠

6 A proof from the Book?

I hope you agree that this proof of Theorem 4 is rather beautiful. We took three simple Facts – each of them very unsurprising penny-plain Facts (the sort of things that should strike you as true after even the briefest, arm-waving, explanation). Then we defined a pair of sets $K/\overline{K}$. Now that did involve a neat little trick, but it's a very familiar type of trick – for it is just same sort of dodge that is involved for example in the Cantorian proof that the powerset of the natural numbers is uncountable. And with these easy ingredients we almost immediately get the result that the truths of arithmetic can't be effectively enumerated.

Now put together Theorem 4 with the elementary result we summed up as Theorem 3, and we get – as desired –

Theorem 5 *If T is a nice, sound theory of arithmetic, then there are arithmetical sentences φ such that T proves neither φ nor $\neg\varphi$.*

So, overall, that is all remarkably neat and simple. Surely this proof of Gödel's Theorem deserves to be in Erdős's Book.

7 What does our proof teach us?

Now, different proofs of a theorem typically highlight different relationships in the local mathematical landscape: so what is revealed by *this* proof?

Perhaps the major lesson is a negative one. Let's say that a Gödel sentence is an arithmetical sentence which – in virtue of some background numerical coding scheme – says of itself 'I am unprovable'. Then our argument shows that a proof of incompleteness *doesn't* have to depend on explicitly constructing a Gödel sentence. Now, Gödel's original paper does go by that route. And he commented himself on analogies between his construction and self-referential paradoxes like the Liar Paradox. That makes too many people, when they first encounter Gödel's argument, feel that there must be something suspicious going on (as if they are being given a paradox rather than a theorem). And even when they get over that first reaction, the impression can remain that Gödel sentences must be some kind of perverse *oddity.* The thought might be encouraged: just as, outside the logic classroom, we get by perfectly well with the idea of truth, without having to worry about sentences like 'this sentence is false', so arithmeticians can surely get by perfectly well without stumbling over sentences that somehow say 'I am unprovable'.

However, our proof shows that the incompleteness phenomenon *isn't* (so to speak) a little local difficulty. You might have hoped that we could effectively list off the supposedly dodgy self-referential Gödel truths, put them to one side, and then give a nice axiomatized formal theory T covering the *rest* of arithmetic, the non-dodgy, sensible stuff. But we can't do that because then we'd then have two effectively enumerable lists, a list of the peculiar unprovable Gödel truths and then the T-theorems. And interweaving these lists would give us an effective enumeration of all the arithmetical truths, which we've just shown is impossible.

What about positive implications of our proof? Well, I'm not sure there are positive lessons that are particularly distinctive: the positive implications are pretty much the same as those of other kinds of proof of Gödel's Theorem. But still, maybe our arguments do make the implications especially vivid. For what we've shown is that any enumerable set of theorems can't include every truth of the type $\neg\exists\mathsf{s}\mathsf{R}(\mathsf{s},\overline{\mathsf{n}})$ which correlates to membership facts about $\overline{K}$. These truths are in a way very simple ones: they are equivalent to truths of the form $\forall\mathsf{s}\neg\mathsf{R}(\mathsf{s},\overline{\mathsf{n}}))$ – which makes them, in the logician's jargon, Π_1 truths. That is to say, they are universally quantified claims about decidable properties. We can easily prove any instance particular $\neg\mathsf{R}(\overline{\mathsf{m}},\overline{\mathsf{n}})$ is true when it is true: yet any properly axiomatized theory will fail to prove some of their true universal quantifications. Add more proof methods, assume new axioms, and we'll be able to prove more and more of those Π_1 truths; but we'll never be able to capture all of them so long as our theory remains under formal control and properly axiomatized.

To borrow a well-known remark from Michael Dummett:

> Gödel's discovery amounted to the demonstration that the class of [principles for establishing quantified claims] cannot be specified exactly once and for all, but must be acknowledged to be an indefinitely extensible class.

Dummett tries to find some deep philosophical significance in this observation, but that's another story, which I don't understand and so don't have anything to say about. But whatever its significance for philosophers, the message for mathematicians is rather cheering. For it means that there will always be a role for mathematicians to dream up and justify new axioms (for example in the higher reaches of set theory), and then new arithmetical truths will become provable. But we'll never complete the job: even if they only care about proving arithmetic truths, mathematicians will never be out of a job.