15 Hilbert and the Second Theorem

Time at last for the Second Theorem. The basic idea is straightforward, the devil is in the details of a full proof (more about that in the next chapter).

Now, we didn't have to do much work to explain why the First Theorem is so interesting – how odd and unexpected to find that even the truths of basic arithmetic escape being completely pinned down in a nicely axiomatized theory! But we do need to do rather more scene-setting to help bring out the significance of the Second Theorem. So in this chapter we also give a cartoon sketch of some history.

15.1 Defining Con_T

When talking about various axiomatized theories T in earlier chapters, we didn't specify a particular formulation of the deductive system built into T (other than that the resulting theory counts as p.r. axiomatized). The logic may or may not have a built-in *absurdity constant* like the conventional '$\bot$'. So henceforth, let's use the absurdity sign in the following way:

Defn. 56. *'$\bot$' is T's built-in absurdity constant if it has one, or else it is an abbreviation for '$\mathsf{0} = \overline{\mathsf{1}}$'.*

Assuming T contains Q, T of course proves $\mathsf{0} \neq \overline{\mathsf{1}}$. So on either reading of '$\bot$', if T also proves $\bot$, it is inconsistent. And conversely, if T's has a standard classical (or indeed intuitionistic) logic and T is inconsistent, then on either reading it will prove $\bot$.

Assuming again that T contains the language of basic arithmetic, and we have a normal Gödel-numbering scheme in place, then (as we now know) we can express provability-in-T in a canonical way by an open T-wff Prov_T, such that $\mathsf{Prov}_T(\overline{\ulcorner \varphi \urcorner})$ is true exactly when φ is a T-theorem.

Hence the wff $\neg\mathsf{Prov}_T(\overline{\ulcorner \bot \urcorner})$ is true if and only if T *doesn't* prove $\bot$, hence (given what we've just said) is true if and only if T is consistent. That evidently motivates the definition

Defn. 57. Con_T *abbreviates* $\neg\mathsf{Prov}_T(\overline{\ulcorner \bot \urcorner})$.

For obvious reasons, the arithmetic sentence Con_T is called a (canonical) *consistency sentence* for T.[1] Since Prov_T is Σ_1, Con_T is Π_1.

Let's have an example for future use. Take your favourite nicely axiomatized set theory – as it might be, Zermelo-Fraenkel theory plus the Axiom of Choice, the standard theory ZFC. We can define the language of basic arithmetic in the language of set theory. So ZFC can itself express provability-in-ZFC using a wff of the theory, $\mathsf{Prov}_{\mathsf{ZFC}}$, and from that we can form the consistency sentence $\mathsf{Con}_{\mathsf{ZFC}}$ which is true iff ZFC is consistent; moreover, this sentence will be an arithmetic sentence (meaning a ZFC version of a sentence of basic arithmetic).

15.2 The Formalized First Theorem

Back, for a moment, to the First Incompleteness Theorem. Assume as usual that we are dealing with a theory T which is p.r. axiomatized and contains Q. Then one half of the First Theorem tells us that

(1) If T is consistent, then G_T is not provable in T.

And given what we said in the previous section, we in fact have a natural way of expressing (1) *inside the formal theory T itself*, i.e. by the conditional

(2) $\mathsf{Con}_T \to \neg\mathsf{Prov}_T(\overline{\ulcorner \mathsf{G}_T \urcorner})$.

Call this wff the *Formalized First Theorem*.

Now let's reflect that, once we have made the cunning move of constructing G_T, the informal reasoning for the First Theorem is in fact *very* elementary. We certainly needed no higher mathematics at all, just relatively straightforward reasoning about arithmetical codings. So we might well expect that if T can reason about arithmetic sufficiently well, *it should itself be able to reflect that elementary reasoning*, and hence itself *prove* our formalized version of the statement of the First Theorem .

In short, we will hope to have the following key result:

Theorem 57. *If T is p.r. axiomatized and contains enough arithmetic, then* $T \vdash \mathsf{Con}_T \to \neg\mathsf{Prov}_T(\overline{\ulcorner \mathsf{G}_T \urcorner})$.

Of course, we'll want to pin down what counts as 'enough arithmetic'; but we'll do that in the next chapter. But here's an initial thought. Maybe T will need to be stronger that Q, because it will presumably need to be able to use induction in generalizing about arithmetized syntax; but containing PA ought to be enough.

[1] There are alternatives for defining consistency sentences. Suppose the wff $\mathsf{Contr}(\mathsf{x},\mathsf{y})$ captures the p.r. relation which holds between two numbers when they code for a contradictory pair of sentences, i.e. one codes for some sentence φ and the other for $\neg\varphi$. Then we could define Con^*_T to be short for the sentence $\neg\exists\mathsf{x}\exists\mathsf{y}(\mathsf{Prov}_T(\mathsf{x}) \wedge \mathsf{Prov}_T(\mathsf{y}) \wedge \mathsf{Contr}(\mathsf{x},\mathsf{y}))$, which says that we can't find two T-provable wffs which contradict each other. This would be another natural way of expressing T's consistency. But, on modest assumptions, Con^*_T is provably equivalent to Con_T: so we'll stick to our standard definition.

15.3 From the Formalized First Theorem to the Second Theorem

Assume we have established Theorem 57 (we'll discuss its proof in the next chapter). Then, if T is p.r. axiomatised, and contains enough arithmetic,

1. $T \vdash \mathsf{Con}_T \rightarrow \neg\mathsf{Prov}_T(\overline{\ulcorner \mathsf{G}_T \urcorner})$

But we know from Theorem 50 that for a p.r. axiomatized T which contains as little arithmetic as Q,

2. $T \vdash \mathsf{G}_T \leftrightarrow \neg\mathsf{Prov}_T(\overline{\ulcorner \mathsf{G}_T \urcorner})$.

Hence from (1) and (2) we have

3. $T \vdash \mathsf{Con}_T \rightarrow \mathsf{G}_T$.

We can therefore immediately infer that

4. If $T \vdash \mathsf{Con}_T$, then $T \vdash \mathsf{G}_T$.

But we know from the First Theorem that

5. If T is consistent, $T \nvdash \mathsf{G}_T$.

So, making the conditions on T explicit again, from (1) we get a (somewhat unspecific) version of Gödel's *Second Incompleteness Theorem*.

Theorem 58. *Suppose T is p.r. axiomatized and contains enough arithmetic: then, if T is consistent, $T \nvdash \mathsf{Con}_T$.*

15.4 The impact of the Second Theorem?

If T could have proved Con_T, that would have been no special evidence *for* T's consistency. After all, an inconsistent theory T can prove anything and therefore prove Con_T in particular.

On the other hand, T's failure to prove Con_T is no evidence *against* T's consistency. We have, hopefully, just found another true Π_1 sentence that T cannot prove, to set alongside G_T.

Hence – you might reasonably suppose – the non-derivability of a canonical statement of T's consistency inside T itself does not show us a great deal.

But that's too fast. For consider this obvious corollary of the Second Theorem:

Theorem 59. *Suppose S is a consistent theory, strong enough for the Second Theorem to apply to it, and W is a weaker fragment of S, then $W \nvdash \mathsf{Con}_S$.*

That's because, if S is strong enough for the Second Theorem to apply to it, it can't prove Con_S. So then, a fortiori, *part* of S can't prove Con_S either.

So, for example, we *can't* take some strong theory like ZFC as the theory S and show that it is consistent by (i) using arithmetic coding for talking about its proofs and then (ii) using uncontentious reasoning already available in some

relatively weak, perhaps purely arithmetical, theory W to show $\mathsf{Con}_{\mathsf{ZFC}}$. The stronger theorem ZFC can't prove $\mathsf{Con}_{\mathsf{ZFC}}$ (assuming it is consistent); so the weaker arithmetic theory W can't prove $\mathsf{Con}_{\mathsf{ZFC}}$ either. And *this* is an important result. Why? We need to fill in some historical background.

15.5 Formalization, finitary reasoning, and Hilbert

Think yourself back to the situation in mathematics a bit over a century ago. Classical analysis – the theory of differentiation and integration – has, supposedly, been put on firm foundations. We have, for example, done away with obscure talk about infinitesimals; and we have traded in an intuitive grasp of the continuum of real numbers for the idea of reals defined as 'Dedekind cuts' on the rationals or 'Cauchy sequences' of rationals. The key idea we've used in our constructions is the idea of a *set* of numbers. And we've been very free and easy with that, allowing ourselves to talk of arbitrary sets of numbers, even when there is no statable rule for collecting the numbers into the set.

This freedom to allow ourselves to talk of arbitrarily constructed sets is just one aspect of the increasing freedom that mathematicians have allowed themselves over the second half of the nineteenth century. We have loosed ourselves from the assumption that mathematics should be tied to the description of nature: as Morris Kline puts it, "after about 1850, the view that mathematics can introduce and deal with arbitrary concepts and theories that do not have any immediate physical interpretation ... gained acceptance". And Cantor could write "Mathematics is entirely free in its development and its concepts are restricted only by the necessity of being non-contradictory".

It is rather bad news, then, if all this play with freely created concepts, and in particular the fundamental notion of arbitrary sets, in fact gets us embroiled in contradiction – as seems to be the case once the set-theoretic paradoxes (like Russell's Paradox) are discovered. What to do?

We might – rather artificially – distinguish two lines of responses to the paradoxes that threaten Cantor's paradise where mathematicians can play freely, which we might suggestively call the *foundationalist* and the *more careful mathematics* responses.

Foundationalist responses to paradox Consider first the option of seeking external "foundations" for mathematics.

We could, perhaps, seek to "re-ground" mathematics by confining ourselves again to applicable mathematics which has, as we would anachronistically put it, a model in the natural world so *must* be consistent.

The trouble with this idea is we're none too clear what this re-grounding in the world would involve – for remember, we are thinking back at the beginning of the twentieth century, as relativity and quantum mechanics are emerging, and any Newtonian confidence that we had about the real structure of the natural world is being shaken.

So maybe we have to put the option of anchoring mathematics in the phys-

ical world aside. But perhaps we could try to ensure that our mathematical constructions are grounded in the mental world, in mental constructions that we can perform and have a secure epistemic access to. This idea leads us to Brouwer's intuitionism. But it depends on an obscure notion of mental construction, and in any case – in its most worked out form – this approach *cripples* a lot of classical mathematics that we thought was unproblematically in good order, rather than giving it a foundation.

What other foundational line might we take? The notion of set that we have used (and seemingly abused) is arguably in some sense a logical notion; perhaps we have got into trouble by going beyond its logical core. How about trying to nail down some incontrovertible logical principles and then find definitions of mathematical notions in purely logical terms, so constraining mathematics to be what we can reconstruct on a firm *logical* footing.

This *logicist* line which we briefly met back in §1.5 has its attractions but it is problematic in various ways. For remember, we are pretending to be situated a hundred or so years back, and at least at this point – leaving aside its beginnings in the as-yet-hardly-read work of Frege – modern logic itself isn't in as good a shape as most of the mathematics we are supposedly going to use it to ground (and indeed what might *count* as incontrovertible logic is still pretty obscure). Moreover, as C. S. Peirce saw, it looks as if we are going to need to appeal to mathematically developed ideas in order to develop logic itself; indeed Peirce himself thought that all formal logic is merely mathematics applied to logic.

Still, perhaps we shouldn't give up yet. The proof is in the pudding: let's see if we can actually do the job and reconstruct mathematics on a logical basis as Frege tried to do (he failed, getting entangled in paradox), and let's write a *Principia Mathematica* ...!

'Mathematical' responses to paradox Maybe, however, back at the beginning of the twentieth century, we just shouldn't be seeking to give mathematics a prior "foundation" after all. Consider: the paradoxes arise within mathematics, and to avoid them (the working mathematican might reasonably think) we just need to do the mathematics more carefully. As Peirce – for example – held, mathematics risks being radically distorted if we seek to make it answerable to some outside considerations. We don't need to look outside mathematics (to the world, to mental constructions, or even to logic) for a justification that will guarantee consistency. Perhaps we just need to improve our mathematical practice, in particular by improving the explicitness of our regimentations of mathematical arguments, to reveal the principles we actually use in 'ordinary' mathematics, and to see where the fatal mis-steps must be occurring when we over-stretch these principles in ways that lead to paradox.

How do we improve explicitness? A first step will be to work towards regimenting the principles that we actually need in mathematics into something approaching the ideal form of an effectively axiomatized formal theory. This is what Zermelo aimed to do in axiomatizing set theory: to locate the principles actually needed for the seemingly 'safe' mathematical constructions needed in

grounding classical analysis and other familiar mathematical practice. And when the job is done it seems that *these* principles don't in fact allow the familiar reasoning leading to Russell's Paradox or other set-theoretic paradoxes. So like the foundationalist/logicist camp, the 'do maths better' camp are also keen on battering a mathematical theory T into a nice tidy axiomatized format and sharply defining the rules of the game: but the purpose of the axiomatization is quite significantly different. The aim now is not to expose foundations, but just to put our theory into a form that – hopefully – we can now show is indeed consistent and trouble-free.

For note the crucial insight of Hilbert's, which we can sum up like this:

> The axiomatic formalization of a mathematical theory T (whether about sets, widgets, or other whatnots), gives us *new* objects (beyond the sets, widgets, or other whatnots) that are *themselves* apt topics for new mathematical investigations – namely the T-wffs and T-proofs that make up the theory!
>
> And, crucially, when we go metatheoretical like this and move from thinking about *sets* (as it might be) to thinking about the syntactic properties of *formalized-theories-about-sets*, we can be moving from considering suites of *infinite* objects to considering suites of *finite* formal objects (the wffs, and the finite sequences of wffs that form proofs). This means that we might then hope to bring to bear, at the metatheoretical level, entirely 'safe', merely *finitary*, reasoning about these suites of finite formal objects in order to prove the consistency of (say) set theory.

Of course, it is a very moot point what exactly constitutes such ultra-'safe' finitary reasoning. But still, it certainly looks as if we will – for instance – need much, much, less than full set theory to reason (not about *sets* but) about a formalized *theory* of sets as a suite of finite syntactic objects. So we might, in particular, hope with Hilbert – in the do-mathematics-better camp – to be able to use a safe uncontentious fragment of finitary mathematics to prove that our wildly infinitary set theory (ZFC, perhaps) is at least syntactically consistent.

So, at this point (I hope!) you can begin to see the attractions of what's called *Hilbert's Programme* – i.e. the programme of showing various systems of carefully reconstructed infinitary mathematics are contradiction-free by giving consistency proofs using safe, finitary, reasoning about the systems considered as formal objects.

But now, enter Gödel, wielding his Second Theorem. He shows that a theory which contains enough arithmetic can't prove its own consistency, let alone the consistency of a stronger theory. For example, we can't use a weak theory W – arithmetic, for example – to prove ZFC consistent by proving $\mathsf{Con}_{\mathsf{ZFC}}$: even full-powered ZFC itself can't prove $\mathsf{Con}_{\mathsf{ZFC}}$. Similarly for other strong theories. Which means that the Second Theorem – at least at first blush – sabotages Hilbert's Programme with its plan of showing various stronger theories are consistent by using a weaker 'finitistic' framework to argue about their syntactic properties.

Famously, when Hilbert heard of the result at a conference where Gödel first announced it, he was not well pleased ....