8

WITTGENSTEIN'S REMARKS ON GÖDEL'S THEOREM

Graham Priest

1. Introduction

Wittgenstein's *Remarks on the Foundations of Mathematics* received perhaps the most lukewarm reception of all of his posthumously published work. For example, Anderson said that it is 'hard to avoid the conclusion that Wittgenstein failed to understand clearly the problems with which workers in the foundations of mathematics have been concerned' (1964: 489); Kreisel called it 'a surprisingly insignificant product of a sparkling mind' (1959: 158); and even Dummett, who is a good deal more sympathetic, after reminding us wisely that these are remarks culled by editors from notebooks that were never intended for publication, averred

> many of the thoughts are expressed in a manner which the author recognized as inaccurate or obscure; some passages contradict others; some are quite inconclusive; some raise objections to ideas which Wittgenstein held or had held which are not clearly stated in the volume.
> (1964: 491)

The remarks on Gödel's theorem, in particular, drew very negative comments. Kreisel thought that Wittgenstein's 'arguments are wild' (1959: 153); Anderson said that they 'indicate that Wittgenstein misunderstood both the content of and the motivation for … Gödel's theorem' (1964: 485); and here is Dummett again: 'other passages again, particularly those on consistency and Gödel's theorem, are of poor quality or contain definite errors' (ibid.).[1] The aim of this paper is to revisit the issue, some half a century on, to see whether these harsh words are justified.

Wittgenstein's remarks on Gödel's (first incompleteness) theorem are contained almost entirely within an appendix of some twenty remarks to Part I of the *Remarks*. The editors tell us (1978: 30) that Part I is based on a typescript that Wittgenstein had intended at one time as a second part of what was to be the *Philosophical Investigations*, but that the text in the appendix was separated from the main body of the material. The appendix is not self-contained, since it alludes to other themes of the *Remarks*, and indeed of the *Investigations*, but its isolation means that it can be considered in a relatively self-standing way. Because of Wittgenstein's writing style in this period, one always has to work hard to determine what is going on. The appendix in question poses this problem *in extremis*. Even seasoned Wittgenstein-interpreters have problems decoding the gnomic utterances. The points of some of the individual paragraphs are difficult to discern; the connections between many of them even more so. Sometimes Wittgenstein is arguing with his imagined interlocutor; sometimes he seems to be wrestling with himself. Yet if we are to give Wittgenstein a fair hearing on the matter, it is essential to understand what, exactly, his view is. So I intend to proceed via a close reading of, and textual commentary on, the passage in question – something that, as far as I know, no one has yet attempted. To this I now turn.

2. The approach to the issue

Wittgenstein's reflections on Gödel's theorem start at some apparent distance from the matter, by observing that not everything written as an indicative sentence has propositional content.[2]

> 1 It is easy to think of a language in which there is not a form for questions, or commands, but question and command are expressed in the form of statements, e.g. in forms corresponding to our: 'I should like to know if …' and 'My wish is that …'
>
> No one would say of a question (e.g. whether it is raining outside) that it was true or false. Of course it is English to say so of such a sentence as 'I want to know whether …' But suppose this form were always used instead of the question? –

He then continues:

> 2 The great majority of sentences that we speak, write and read are statement sentences.
>
> And – you say – these sentences are true or false. Or, as I might also say, the game of truth-functions is played with them. For assertion is not something that gets added to the proposition, but an essential feature of the game we play with it. Comparable, say, to the characteristic of chess by which there is winning and losing in it, the winner being the one who takes the other's king. Of course, there could be a game in a certain sense very near akin to chess, consisting in making the chess moves, but without there being any winning or losing in it; or with different conditions of winning.

> 3 Imagine it were said: A command consists of a proposal ('assumption') and the commanding of the thing proposed.

> 4 Might we not do arithmetic without ever having the idea of uttering arithmetical *propositions*, and without ever having been struck by the similarity between a multiplication and a proposition?
>
> Should we not shake our heads, though, when someone shewed us a multiplication done wrong, as we do when someone tells us it is raining, if it is not raining? – Yes; and here is a point of connexion. But we also make gestures to stop our dog, e.g. when he behaves as we do not wish.
>
> We are used to saying '2 times 2 is 4', and the verb 'is' makes this into a proposition, and apparently establishes a close kinship with everything we call a 'proposition'. Whereas it is a matter only of a very superficial relationship.

Though these remarks raise a number of different matters, let us pass them over. The important thing for now is that by the time we get to remark 4, it is clear that Wittgenstein is considering the idea that arithmetic equations, though they may be written as indicative sentences, like sentences such as 'I want to know whether it is raining', do not really have propositional content. This is a theme in Wittgenstein that goes all the way back to the *Tractatus*. (See, e.g., *Tractatus* 4.46ff., 6.2ff.). There, statements of logic and mathematics are argued to be *unsinnig*, to carry no information content at all. Though Wittgenstein has long since given up the views of the *Tractatus* by this point, he is still playing with thought that statements of mathematics and logic have no content. This is the context which triggers his ruminations on Gödel's theorem. In case there is any doubt at all about this matter, he returns explicitly to the subject in his final remark (20): 'Here one needs to remember that the propositions of logic are so constructed as to have *no* application as *information* in practice. So it could very well be said that they were not *propositions* at all …'[3]

3. The posing of the problem and an initial solution

Against this background, the next remark introduces the subject of Gödel's theorem.

> 5 Are there true propositions in Russell's system, which cannot be proved in his system? – What is called a true proposition in Russell's system, then?

It would seem that the theorem – or at least, one way that it is often phrased – is being raised as an objection to the view of mathematics in question. If it is right, mathematical statements, or at least some of them, must have propositional content, indeed true content. But what, asks Wittgenstein, does truth mean in this context? The next paragraph answers the question.

> 6 For what does a proposition's '*being true*' mean? '*p*' *is true* = *p*. (That is the answer.)
>
> So we want to ask something like: under what circumstances do we assert a proposition? Or: how is the assertion of the proposition used in the language-game? And the 'assertion of the proposition' is here contrasted with the utterance of the sentence e.g. as practice in elocution, – or as *part of* another proposition, and so on.
>
> If, then, we ask in this sense: 'Under what circumstances is a proposition asserted in Russell's game?' the answer is: at the end of one of his proofs, or as a 'fundamental law' (Pp). There is no other way in this system of employing asserted propositions in Russell's symbolism.

Wittgenstein invokes a redundancy theory of truth. To say that *p* is true is to say no more and no less than *p* itself. The position is familiar from elsewhere in the later writings (e.g. *Philosophical Investigations*, remark 136). It is clearly a substantial account of truth, but it is not, of course, the model-theoretic account of truth that would normally be invoked by modern logicians in a discussion of Gödel's theorem. Let us not try to adjudicate the point here, but just note it.

Given a redundancy account of truth, the pertinent question then becomes under what conditions we are prepared to assert *p*. Wittgenstein is obviously not supposing that asserting a sentence means that it has a propositional content. He contrasts asserting, instead, with other kinds of utterance, e.g. elocutory. Then there emerges another familiar Wittgensteinian theme. Assertion is relative to a language game. That is, when we employ sentences expressed in the language of *Principia*, there are rules that determine when a sentence may or not be asserted. In particular, thinks Wittgenstein, such a sentence may be asserted when it occurs at the last line of a *Principia* proof (maybe a one-line proof). In other words, to be a true *Principia* sentence is just to be provable in the *Principia* axiom system. Whatever one thinks about a redundancy theory of truth, and of the theory of language games, this fact at least gives someone who does not subscribe to these views a way of understanding Wittgenstein in their own terms. When they hear Wittgenstein talk of a truth of *Principia*, they can simply hear him as meaning 'provable in the *Principia* axiom system'. Gödel's theorem of course applies to other languages and other axioms systems. Doubtless, Wittgenstein was aware of this fact. Doubtless, also, he would have taken his remarks to apply equally to any similar system. He treats *Principia* simply as an example. So shall I.

It should be noted that Wittgenstein concedes that the sentences of *Principia* are true/false in an appropriate sense. But the sense is only that of being generated by certain rules. This is quite compatible with their having no propositional or information content, as Wittgenstein is concerned to defend. But Wittgenstein's identification of truth and provability itself seems to be controverted by Gödel's theorem, as the interlocutor is quick to point out.

> 7 'But may there not be true propositions which are written in this symbolism, but are not provable in Russell's system?' – 'True propositions', hence propositions which are true in *another system*, i.e., can rightly be asserted in another game. Certainly; why should there not be such propositions; or rather: why should not propositions – of physics, e.g. – be written in Russell's symbolism? The question is quite analogous to: Can there be true propositions in the language of Euclid, which are not provable in his system, but are also true? – Why, there are even propositions which are provable in Euclid's system, but are *false* in another system. May not triangles be – in another system – similar (*very* similar) which do not have equal angles? – 'But that's just a joke! For in that case they are not 'similar' to one another in the same sense' – Of course not; and a proposition which cannot be proved in Russell's system is 'true' or 'false' in a different sense from a proposition of *Principia Mathematica*.

Wittgenstein, fortified by his understanding of truth, makes the obvious reply. Of course there can be sentences in the language of *Principia* that are not provable in the *Principia* axiom system, but that are provable in another axiom system. This reply would be comfortable to any modern logician. The Gödel undecidable sentence is not provable in the theory itself, but is provable (or can be proved to be true) in a metalanguage/metatheory.

Wittgenstein points out, correctly, that this is similar, e.g., to certain sentences being provable in Euclidean geometry but not some other geometry. The interlocutor takes this to be a joke: meanings change in the geometric case. Again, Wittgenstein replies correctly: meaning changes in this case too. This time it is 'provable' that is ambiguous; the undecidable sentence is not provable in the *Principia* axiom system, but it is provable in a metatheory.

But maybe this does not get to the heart of the worry. Wittgenstein continues to muse:

> 8 I imagine someone asking my advice; he says: 'I have constructed a proposition (I will use '*P*' to designate it) in Russell's symbolism, and by means of certain definitions and transformations it can be so interpreted that it says: '*P* is not provable in Russell's system'. Must I not say that this proposition on the one hand is true, and on the other hand is unprovable? For suppose it were false; then it is true that it is provable. And that surely cannot be! And if it is proved, then it is proved that it is not provable. Thus, it can only be true but unprovable.'

The key thought here is that the undecidable sentence, *P*, can be interpreted so as to *mean* that it is not provable in *Principia*. Thus:

(I) *P* iff *P* is not provable in *Principia*.[4]

Note this equivalence. It is between a sentence of the *Principia* language and a sentence of the metalanguage. And since these come from different 'language games', it is not at all obvious – from Wittgenstein's perspective – that the two sentences in question mean the same thing. The tenability of the equivalence – and indeed, exactly what its two sides mean – will be cruces of the subsequent discussion.

Given (I) we may reason as follows:

(A) If *P* is false, i.e. its negation is true, then by redundancy account of truth, it is provable. By the soundness of *Principia*, this is impossible; so it is true.
(B) If *P* is proved, then, again by the soundness of *Principia*, it is true, and hence *P* is not provable; so it is not provable.

We have arguments for both the truth of *P* and its unprovability. This directly challenges Wittgenstein's identification of truth with provability, and so raises the spectre that the sentence has a content which is true in some more substantive sense – in which case he may have to concede that it has some real content after all; or, and perhaps worse, if he really does want to insist on the identification of truth with provability, we have a flat contradiction. Note that the remark fails to mention the soundness of *Principia*. But note, also, that if one does identify 'true in *Principia*' with 'provable in *Principia*', as Wittgenstein wants to, both soundness and its converse are true by definition.

In reply, Wittgenstein insists that we must be clear on the ambiguity that he has diagnosed in the notion of proof, i.e. truth. He continues:

> Just as we ask : ' "provable" in what system?', so we must ask: ' "true" in what system?' 'True in Russell's system' means, as was said: proved in Russell's system; and 'false in Russell's system' means the opposite has been proved in Russell's system. – Now what does your 'suppose it is false' mean? *In the Russell sense* it means 'suppose the opposite is proved in Russell's system'; *if that is your assumption*, you will now presumably give up the interpretation that it is unprovable. And by 'this interpretation' I understand the translation into the English sentence. – If you assume that the proposition is provable in Russell's system, that means that it is true *in the Russell sense*, and the interpretation '*P* is not provable' again has to be given up. If you assume that the proposition is true in the Russell sense, *the same* thing follows. Further: if the proposition is supposed to be false in some other than Russell's sense, then it does not contradict this for it to be proved in Russell's system. (What is called 'losing' in chess may constitute winning in another game.)

Suppose that 'false' means that its negation is provable in *Principia*. Then, assuming (I), the reasoning (A) shows that *P* is true, i.e., provable in *Principia*. But in that case, we ought to give up the interpretation (I). After all, the left-hand side would be true, and the right-hand side false. Similarly, if 'true'/'provable' means 'provable in *Principia*', (B) shows that *P* is not true/provable. Again, this fact gives us ground to reject (I). If, on the other hand, 'false' means something else – then the fact that something is false does not conflict with its being provable in *Principia*.

The next remark continues:

> 9 For what does it mean to say that *P* and '*P* is unprovable' are the same proposition? It means that these two English sentences have a single expression in such-and-such a notation.

The remark would appear to justify glossing the claim about *P* being interpretable as '*P* is unprovable' simply as meaning that we can use the latter sentence anywhere we use the former, and vice versa. There is a small slip here. Wittgenstein seems to have forgotten that *P* is not a sentence of English, but of *Principia*ese. But the upshot of the thought seems unobjectionable enough.

At any rate, the thrust of Wittgenstein's thought on the matter is clear: contradiction threatens to arise only if one endorses both a certain notion of truth/provability and the equivalence (I) for this notion. But this equivalence might well be legitimately resistible for such a notion. Note that the contemporary orthodoxy would be to endorse (I), or something like it. Given that the proof predicate, $B(x,y)$, really represents provability, *P* – that is, $\neg\exists x B(x,\mathrm{n})$, whose code is n, with numeral n – is true *in the standard* model iff *P* is not provable. Wittgenstein, however, is operating with a different notion of truth, and so this argument for (I) is not open to him.

4. Countenancing inconsistency

But now Wittgenstein continues to muse:

> 10 'But surely *P* cannot be provable, for, supposing it were proved, then the proposition that it is not provable would be proved.' But if this were now proved, or if I believed – perhaps through an error – that I had proved it, why should I not let the proof stand and say I must withdraw my interpretation '*unprovable*'?

He reiterates argument (B) to the effect that *P* is not provable, and then asks the crucial question. Suppose that (B) constitutes a proof, or at any rate, that I believe that it is does, *why* should I not let the proof *and* the interpretation (I) stand?[5] In this case, I would have proved *P* in *Principia*, but also proved that it is not provable in *Principia*. We have a contradiction. But so what?[6]

> 11 Let us suppose I prove the unprovability (in Russell's system) of *P*; then by this proof I have proved *P*. Now if this proof were one in

Russell's system – I should in that case have proved at once that it belonged and did not belong to Russell's system. – That is what comes of making up such sentences. – But there is a contradiction here! – Well, then there is a contradiction here. Does it do any harm here?

12 Is there any harm in the contradiction that arises when someone says: 'I am lying. – So I am not lying. – So I am lying. – etc.'? I mean: does it make our language less usable if in this case, according to the ordinary rules, a proposition yields its contradictory, and vice versa? – the proposition *itself* is unusable, and these inferences equally; but why should they not be made? – It is a profitless performance! – It is a language-game with some similarity to the game of thumb-catching.

13 Such a contradiction is of interest only because it has tormented people, and because this shews both how tormenting problems can grow out of language, and what kind of things can torment us.

It is perhaps these remarks which have drawn the ire of commentators more than any others. Wittgenstein countenances the possibility that it has been shown that *P* both is and is not provable in *Principia* – and even that this might be proved in *Principia* itself; and he seems happy with this idea. This is guaranteed to touch a raw nerve in most, the superstitious dread of contradictions, as Wittgenstein himself puts it in remark 17. Wittgenstein's apparent preparedness to accept contradictions is not isolated. At other places in the *Remarks* and elsewhere, Wittgenstein dallies with contradiction.[7] He claims that contradictions of the kind we have here are useless, that drawing them is pointless, but that since they do not have any impact on the rest of our language, they do no harm. If he had enforced the doctrine that meaning is use, he might have gone on to claim that contradictions of this kind are meaningless. And indeed, at certain times, he was sympathetic to this view.[8] But even in the *Investigations*, he never held that meaning was simply to be equated with use. (See, e.g., remark 43.)

Wittgenstein also observes the similarity between the contradiction in question and contradictions of the liar variety. Indeed, the paradox in question can well be seen as a paradox of self-reference of the same kind. Consider the sentence *A*, of the form '<*A*> is not provable' – this sentence is not provable – angle brackets represent some naming device. Here, provability is to be understood in the naive sense of being demonstrated by some argument or other. If *A* is provable, then, since what is provable is true, *A* is true; so <*A*> is not provable. Hence, <*A*> is not provable. But we have just proved this; that is, <*A*> is provable. This is a version of the 'Knower paradox'. Sometimes it is called 'Gödel's paradox'.[9] In fact, if one identifies truth with provability, as does Wittgenstein, Gödel's paradox and the liar collapse into each other.

At any rate, how might one object to Wittgenstein's preparedness to accept contradiction? There are several ways. First, of course, one might object that

Wittgenstein's view cannot be correct since such a contradiction flies in the face of the Law of Non-Contradiction: contradictions cannot be true. Wittgenstein would doubtlessly be unimpressed by this logical shibboleth – and given his other views, rightly so. To be true in a language game is simply to be produced by correctly following the rules of that game. And if there is a game that correctly generates sentences of the form A and $\neg A$, so be it. Contradictions are true in that language game. Maybe the language game that we play in English about proof in *Principia* is like that.

Perhaps more pointedly, one might object that even if the contradiction is true, it is hardly harmless as Wittgenstein claims. It ruins the language game altogether. For from it, by well-known rules of the game, we can derive every sentence: contradictions entail everything. It is unlikely that Wittgenstein would have been much impressed by this either. He would doubtless point out that, as a matter of fact, we do not infer everything from a liar-type contradiction. An objector might say that this is beside the point: whether we do or not, the rules of the game allow us to do so, so the game is useless. But here, again, Wittgenstein would probably object. What are the rules of the game that govern inference about proof in *Principia*? The objector is just *assuming* that these are the rules of something like classical logic, in which contradictions entail everything. Doubtless, such a point would not have impressed had it been made at the time of the publication of the *Remarks*. Most people then could see little alternative to the rules of 'classical logic' – or, at least, if there were other logics, such as intuitionist logic, they, too, were explosive.

History has come to Wittgenstein's aid here, though. We now know that there are many paraconsistent logics, logics in which contradictions do not imply everything. Indeed, one of the main motivations for the development of such logics was precisely the thought that the correct logic for reasoning about paradoxes of self-reference is a paraconsistent logic.[10] Wittgenstein, of course, knew nothing of such future developments.[11] But in a remark of great prescience (made in 1930), he foresaw their development: 'Indeed, even at this stage, I predict a time when there will be mathematical investigations of calculi containing contradictions, and people will actually be proud of having emancipated themselves from consistency.'[12]

Thus, he was not at all unsympathetic to the idea of paraconsistent logic; and the objection that the contradiction in question ruins the language, with its rules, since these deliver everything, is adequately met if the underlying logic of the game is paraconsistent. (One might, of course, argue that this particular contradiction is objectionable for some other reason. But that would be a different matter.)

This is not an end of the matter, however; and Wittgenstein is not so quickly off the hook. He is not only considering the possibility that we might prove a contradiction when we reason about what is provable in *Principia*. He is also considering the possibility that the contradiction might be provable in *Principia* itself. Argument (B) is not, or need not be, an argument in *Principia*. It is simply an argument in 'the metalinguistic language game'. However, in remark 11 Wittgenstein explicitly countenances the possibility that this proof, or one like it, can be run in *Principia* itself ('Now if this proof were one in Russell's system …') – though nothing that has gone before seems to force this suggestion on him. (There is perhaps another small slip here. He takes a proof of P and $\neg P$ within *Principia* to show that P is both provable in *Principia* and is not. This does not follow – at least, not without (I) and its contrapositive.)

In this case Wittgenstein's preparedness to accept contradiction is surely mistaken. For, quite explicitly, the rules of inference of *Principia are* those of classical logic. So a contradiction in the system does do great damage. It ensures that everything is 'true' in *Principia*. Certainly, this does not render the system and its contradictions entirely useless: we could still use it to practice calligraphy, or to illustrate what a trivial system is like, etc. But it is useless should we wish to apply it to tell us anything interesting about numbers, in the way that it is normally taken to be applicable.

Even here, however, one may salvage something of importance. *Principia* is just an example of the sort of thing that Wittgenstein is talking about. And it is true that similar considerations apply to any formal system of arithmetic based on an explosive logic. But formal systems of arithmetic can be based on a paraconsistent logic. Such systems are, in fact, now well known – even inconsistent systems of number theory. Such systems may contain all of standard number theory, and even be complete, in the sense that every sentence or its negation (or both) is derivable.[13]

For such systems, not only is it clear, as it is not in the case of our informal metatheoretic reasoning in English about provability, that the logic is paraconsistent; it is also demonstrable that the contradictions are quarantined to within certain areas, and do not destroy the general applicability of the system. Finally, there are systems of this kind where the 'Gödel undecidable sentence' $\neg\exists xB(x,\mathrm{n})$, is such that both it and its negation are provable – just as Wittgenstein envisages here. Thus, such arithmetics can formally encode Gödel's paradox (interpreting provability as provability within the system).[14] Even though Wittgenstein is wrong about *Principia*, then, his view may be quite right when applied to certain paraconsistent formal arithmetics.

Maybe, then, we do not have to give up the equivalence (I). We may have to live with the consequence that a contradiction is true, but perhaps we can do that. The upshot is that there are sentences that are true and false, provable and having a provable negation – whether or not we are talking of truth/provability within a formal system game or within our metatheoretic game. Not even this threatens Wittgenstein's original claim that arithmetic sentences do not express propositions.

5. The question of meaning

Perhaps we can, then, accept a contradiction. But this hardly finishes matters. If we have both proved P and proved that P is unprovable, what on earth does the proof of unprovability *mean*? Wittgenstein essays an obvious possibility.

> 14 A proof of unprovability is as it were a geometrical proof; a proof concerning the geometry of proofs. Quite analogous e.g. to a proof that such-and-such a construction is impossible with ruler and compass. Now such a proof contains an element of prediction, a physical element. For in consequence of such a proof we may say to a man: 'Don't exert yourself to find a construction (of the trisection of an angle, say) – it can be proved that it can't be done'. That is to say: it is essential that the proof of unprovability should be capable of being applied in this way. It must – we might say – be a *forcible reason* for giving up the search for a proof (i.e. for a construction of such-and-such a kind).

A contradiction is unusable as such a prediction.

One might think of an unprovability proof as like a proof that some particular geometric task cannot be performed by ruler and compass. It is, in fact, quite natural to think of it in this way. The geometric proof shows that a certain shape cannot be produced by certain procedures; and an unprovability result might be taken as an indication that a certain syntactic geometric shape cannot be produced by certain procedures. The geometric proof obviously contains an empirical element. We know that no one is, as a matter of fact, going to produce this shape with those procedures, no matter how hard they try. Can one think of the content of the unprovability result in the same way: no one is ever going to produce the shape of *P* if they stick to the rules of the game? No. Assuming (I), the proof of unprovability is itself a proof of *P*. And if we have a proof of *P*, we clearly can't interpret the statement of unprovability in that way! The contradictory nature of what has been proved evacuates it of any empirical content of this kind – or as Wittgenstein has already put it in remark 12, paradoxical sentences are useless.

But if the statement of unprovability is not to be interpreted in this way, what else could it mean? Wittgenstein tries another, and more general, tack.

> 15 Whether something is rightly called the proposition '*X* is unprovable' depends on how we prove this proposition. The proof alone shows what counts as the criterion of unprovability. The proof is part of the system of operations, of the game, in which the proposition is used, and shews us its 'sense'.

Thus the question is whether the 'proof of the unprovability of *P*' is here a forcible reason for the assumption that a proof of *P* will not be found.

> 16 The proposition '*P* is unprovable' has a different sense afterwards – from before it was proved.
>
> If it is proved, then it is the terminal pattern in the proof of unprovability. – If it is unproved, then *what* is to count as a criterion of its truth is not yet *clear*, and – we can say – its sense is still veiled.

To put it simply – always a dangerous and potentially misleading thing when dealing with Wittgenstein – the meaning of a mathematical theorem is determined by its proof. The proof gives us, as it were, a criterion for asserting it. If we have not yet found a proof, we have no grounds for asserting it; and in a sense don't know what it means. And if we find a new proof, we have a new criterion for asserting it, and so, in a sense, its meaning has changed. If this is right, to find out what it means to say that *P* is unprovable in this context, we have to look at the proof of this sentence. The proof that *P* is unprovable will give us the sense of this conclusion – and, presumably, this will not be to the effect that a construction of certain kind cannot be found.

Wittgenstein does not attempt to defend his views concerning the connection between meaning and proof here, but isolated remarks on the connection are scattered throughout the *Remarks* (see, e.g., pp. 162, 367). Together, they constitute one of its more difficult and problematic themes. Let me tease apart some of its aspects.

To determine the sense of a proposition, we must look to its proof. But as remark 15 reminds us, the proof is not to be removed from the whole network of operations that constitute the notion of proof. Hence, it might be more accurate to say that the meaning of a sentence is constituted by its proof conditions, which state such systematic connections. In such a form, the view is familiar enough. For example, as is well known, this is exactly the account of meaning given by intuitionists. Of course, a classical logician may object to this. The sense of a sentence is given by its truth conditions, not its proof conditions. But whatever turns on this disagreement, we may bypass it here. For Wittgenstein has identified truth (in a game) with proof (in that game). Given this assumption, we may therefore talk of proof conditions and truth conditions indifferently. And both classical and intuitionist logicians may agree that the meaning of a sentence is given by its proof/truth conditions.

This is as far as remark 15 takes us. Unfortunately, Wittgenstein goes further in remark 16. It is not just the *possible* proofs (the proof conditions) that determine the meaning of a sentence, but the proofs that are actually in our hands. This thesis has an air of paradox about it. If we don't have a grasp of what a sentence means till we have a proof of it, how could we possibly recognize one when we see it? And if the existence of a new proof changes the meaning of a sentence, why doesn't this show that the old proofs no longer work? After all, they proved the sentence with its old meaning. And the order in which we find the proofs is, presumably, of no logico-semantic significance, so if the new proof undercuts the old proofs, presumably the old proofs undercut the new one. There is a tangle of issue here, and this is not the place to discuss them. In the end, I think, Wittgenstein's position is untenable. Fortunately, then, we can largely bypass the issue. The remarks of Wittgenstein that follow make reference to this stronger view only once, and then not in an essential way. Nor is this an accident. In the case at hand, we actually have the proof of the unprovability of *P*, so we can examine the sense that this delivers to its result. As far as the present discussion goes, we can simply assume that the sense of a sentence is given by its proof (= truth) conditions, and leave Wittgenstein's more extreme view for others to worry about.

Before we move on, one further comment concerning Wittgenstein's extreme verificationism is required. This plausibly entails a certain voluntarism about proof. Suppose that we are about to apply a rule of inference. For the sake of illustration, suppose that we have established *A* and *A* ® *B*, and are about to apply *modus ponens*. This would allow us to infer *B*; but doing so, by giving us a new proof of *B*, would change its meaning; and there is no reason to suppose that *A* and *A* ® *B*, with their present meanings, entail *that*. Of course, once we have applied *modus ponens*, *B* will have changed its meaning. And with the new meaning of *B*, *A* and *A* ® *B* will entail *that*. But we now have a choice. We can decide to apply *modus ponens*, and accept the consequences; or we can decide not to, and hence take the application of the rule to be invalid. Thus, a putative proof presents us with a *decision* as to whether or not to accept it. Wittgenstein certainly seems to endorse such voluntarism at some places (e.g. *Remarks*: 163, 268). But it is clear that it sits ill with the phenomenology of proof. Indeed, the conclusion seems to be in some tension with the *Investigations* account of rule-following. Applying *modus ponens* (or a similar such rule) is simply something that, in the last instance, one does blindly. If I choose not to apply it, then my rule-following peers will simply say that I have not understood. It is, indeed, natural to suppose that Wittgenstein's verificationism, with its attendant voluntarism, is a feature of Wittgenstein's later middle-period thought that was jettisoned by the time of the mature *Investigations*.[15] Again, then, it would be wise not to let too much of substance hang on it.

6. Remark 17

Having spelled out the connection between meaning and proof, Wittgenstein, as one would expect, next applies the idea to the proofs of *P* and its unprovability. The next remark poses one of the more difficult exegetical challenges of the whole passage. I shall therefore break the commentary up into parts. The remark starts as follows:

> 17 Now, how am I to take *P* as having been proved? By a proof of unprovability? Or in some other way? Suppose it is by a proof of unprovability. Now, in order to see *what* has been proved, look at the proof. Perhaps it has here been proved that such-and-such forms of proof do not lead to *P*. – Or, suppose *P* has been proved in a direct way – as I should like to put it – and so in that case there follows the proposition '*P* is unprovable', and it must now come out how this interpretation of the symbols collides with the fact of the proof, and why it has to be given up here.

Wittgenstein draws a distinction between 'proving *P* by a proof of unprovability' and proving it directly. The distinction is not entirely clear, but I take it that this is essentially the distinction between proving that *P* is not provable in the metalanguage and proving *P* in the object language. In the first case, Wittgenstein merely reminds us again to look to the proof to see the sense of its conclusion; he suggests that the sense might be to the effect that such and such forms of proof do not lead to *P*. Presumably, this is like the case of a proof that some construction cannot be effected with ruler and compass. Wittgenstein does not tell us explicitly what conclusions should be drawn from this, but he has already indicated in remark 8 that we should expect to have to give up (I). If (I) holds, then *P* itself would have to mean that there is no proof of it. But if it really is, or were, 'true in Russell's system', *P would* be provable in *Principia*; in which case, it couldn't mean that. To see what it does mean, we would have to look to the *Principia* proof of *P* – if any. At any rate, it would seem that the interpretation (I) can be maintained only by equivocation. Without such equivocation, it has to be given up, in which case, as we have seen, no contradiction threatens.[16]

The other possibility is that *P* is proved within *Principia* itself. This demonstrates that such a proof can be effected. Whatever *P* means, then, it is presumably not something to the effect that there is no such proof. Again, we have to look at the *Principia* proof to establish what, exactly, the result means. But again, as Wittgenstein explicitly says this time, it looks as though we will want to reject the interpretation (I), so no contradiction arises.

The passage goes on to consider the possibility that $\neg P$, i.e. on the assumption (I) (or at any rate its contrapositive), that *P* is provable, is proved. In what does such a proof consist?

> Suppose however that not-*P* is proved. Proved *how*? Say *P*'s being proved directly – for from that follows that it is provable, and hence not-*P*. What am I to say now, '*P*' or 'not-*P*'? Why not both? If someone asks me 'Which is the case, *P*, or not-*P*?' then I reply: *P* stands at the end of a Russellian proof, so you write *P* in the Russellian system; on the other hand, however, it is then provable and this is expressed by not-*P*, but this proposition does not stand at the end of a Russellian proof, and so does not belong to the Russellian system. – When the interpretation '*P* is unprovable' was given to *P*, this proof of *P* was not known, and so one cannot say that *P* says: *this* proof does not exist. – Once the proof has been constructed, this has created a *new situation*: and now we have to decide whether we will call *this* a proof (a *further* proof), or whether we will still call *this* the statement of unprovability.

Suppose that $\neg P$, i.e. that *P* is provable, is demonstrated by producing a *Principia* proof of *P*. In fact, we have just considered that possibility. Its upshot, as we saw, was that *P* and '*P* is not provable' have to be taken to have different senses. We can then say both P and $\neg P$: P is true in the sense that it occurs at the end of a *Principia* proof. But P is provable ($\neg P$), since this is proved in the metalanguage. If there is to be no equivocation, (I) must be given up, and consistency is maintained.

Actually, this is not quite what Wittgenstein says. This is the one place where his radical views of the nature of proof kick in. He claims that when (I) was endorsed, a *Principia* proof of P was not known. (Assuming *Principia* to be consistent, there is none.) Hence, if such a proof were to turn up, P cannot mean that *that* proof does not exist. (This seems implausible. If I claim that all swans are white, and an Australian swan turns up, surely it was part of my claim that that swan was white. The claim is just false.) Moreover, presented with a putative such a proof, we have a choice: we can either accept it as *bona fide*, and interpret the metatheoretic statement as not denying its existence, or insist that it does deny its existence, and reject the putative *Principia* proof. This move depends on Wittgenstein's voluntarism, and I doubt that in the end one can make much sense of it. But no matter; the upshot of this way of looking at things is the same as that which I have just described. If P were to have a *Principia* proof, then the interpretation (I) can be maintained only by giving the right-hand side a different meaning. Inconsistency, then, does not arise.

What of the other possibility, that $\neg P$ is proved directly?

> Suppose not-P is directly proved; it is therefore proved that P can be directly proved! So this is once more a question of interpretation – unless we now also have a direct proof of P. If it were like that, well, that is how it would be.
>
> (The superstitious dread and veneration by mathematicians in face of contradiction.)

The situation is similar to the previous one. Suppose that we were to have a *Principia* proof of ¬P. Assuming (I) (or its contrapositive) we can prove that P is provable. So if P is not directly provable, then, whatever 'P is provable' means it cannot be taken to assert the existence of such a proof. (I) can therefore be endorsed only by giving the right-hand side a different meaning. Of course, if one can directly prove P as well, then this is no longer the case. But in that case, we would be able to prove both P and $\neg P$ in *Principia*, and the system would be inconsistent. Both Wittgenstein and I have already discussed the situation that would then arise.

What is the upshot of all this? Wittgenstein does not tell us which of these situations we are in. He simply covers all bases, including the possibility that *Principia* is inconsistent. In most of the situations the upshot of his reasoning is that one has to reject (I), or, at least, reinterpret one of its sides. If one does this, then, as he already pointed out in remark 8, no contradiction arises. If *Principia* is inconsistent, however, the situation is different. We can maintain (I). We just live with the contradiction. As I already observed in the commentary on remarks

11–13, Wittgenstein's optimism is, strictly, unjustified here.[17] But as regards the inconsistent arithmetics, in which both the Gödel sentence and its negation are provable, he may well be right.

In any case, and whether we reject (I) or accept the contradiction, to return to Wittgenstein's original concern with the matter, the gap between truth (in *Principia*) and provability (in *Principia*) fails to open up, and so the results of Gödel's theorem do not pose an objection to Wittgenstein's identification of the two, or to his original thought that sentences of arithmetic do not express propositions.[18]

7. Final remarks

Most of the philosophical action is now over. The final few remarks mop up. The interlocutor has one last shot.

> 18 'But suppose, now, that the proposition were *false* – and hence provable?' – Why do you call it 'false'? Because you can see a proof? – Or for other reasons? For in that case it doesn't matter. For one can quite well call the Law of Contradiction false, on the grounds that we very often make good sense by answering a question 'Yes and no'. And the same for the proposition '((*p*=*p*' because we employ double negation as a *strengthening* of the negation and not merely as its cancellation.

'Okay,' says the interlocutor, 'I can't say that *P* is true and not provable; but I can't say it's false either. For if it's false, it must be provable, and so not true.' Wittgenstein merely has to remind that just as 'true' and 'provable' are ambiguous, so is 'false'. To say that it is false could mean that there is a *Principia* proof of its negation. Well, we have already dealt with that in the previous remark. If it means something else, then the claim may have a sense that is quite compatible with its being provable in *Principia*. In that case, though '*P* is false and provable' may look like a contradiction it isn't really, since we have a change of sense in the conjuncts. After all, we often make sense of answers like 'yes and no': 'yes' in one sense and 'no' in another. Similarly, if doubling a negation just means strengthening it, then it is clear that ((*p* may not mean the same as *p*.

Wittgenstein then goes on to the attack.

> 19 You say: '…, so *P* is true and unprovable'. That presumably means: 'Therefore *P*'. This is all right with me – but for what purpose do you write down this 'assertion'? (It is as if someone had extracted from certain principles about natural forms and architectural style the idea that on Mount Everest, where no one can live, there belonged a châlet in the Baroque style. And how could you make the truth of the assertion plausible to me, since you can make no use of it except to do these bits of legerdemain?[19]

Whatever one means when one correctly expresses the result of Gödel's theorem by saying that a sentence is true but unprovable, that's fine, as we have seen. But why bother in the first place? Until you have shown some independent use for the assertion, you haven't succeeded in saying anything much. As remark 12 noted, the claim seems quite useless. And if the claim has no content, it cannot pose a challenge to Wittgenstein's original view about the nature of mathematical sentences.

Finally, Wittgenstein returns to the original issue explicitly, bringing the last thought to bear on it.

> 20 Here one needs to remember that the propositions of logic are so constructed as to have *no* application as *information* in practice. So it could very well be said that they were not *propositions* at all; and one's writing them down at all stands in need of justification. Now if we append to these 'propositions' a further sentence-like structure of another kind, then we are all the more in the dark about what kind of application this system of sign-combinations is supposed to have; for the mere *ring of a sentence* is not enough to give these connexions of signs any meaning.

Mathematical sentences do not have propositional content: they contain no information. And if the sentence P has no content, this would appear to go for 'P and P is not provable' in spades.

8. Conclusions

Let me now draw the threads of the discussion together and point out its most significant features. The context for Wittgenstein's discussion of Gödel's theorem is the thought that sentences of mathematics have no propositional content. Wittgenstein is concerned to defend the idea that this might be the case. Gödel's result is introduced as an objection to this view – and especially, as it turns out, to Wittgenstein's identification of truth with provability. By the end of the remarks, the objection has been disposed of, and the view still stands.

Wittgenstein discusses Gödel's result against two background assumptions. The first is a redundancy theory of truth, and the second is the theory of language games. The second of these is standard Wittgensteinian fare. The first is less so, but views of this kind now have greater popularity that they did when Wittgenstein wrote the *Remarks*. At any rate, neither is a silly view.

These assumptions are not the standard ones that are made in contemporary discussions of Gödel's result. In particular, most such discussions would invoke the model-theoretic account of truth. Did Wittgenstein know about such a possibility? Did he understand it? Who knows? The text is simply silent on the matter. However, the orthodox reaction to Gödel's result is to insist on the distinction between object- and metalanguage. A certain sentence cannot be proved in the object-language, but it can be proved (to be true) in the metalanguage. In effect, Wittgenstein gives this reply soon after raising the matter. Assertions to the effect that Wittgenstein *mis*understood Gödel's theorem therefore seem misplaced.

According to the model-theoretic account of truth, the equivalence (I) is unproblematic. In the context that Wittgenstein is operating in, it is not, and this allows him to question it. In particular, he can ask exactly what the right-hand side means. This allows him to take the discussion into areas beyond those normally countenanced in discussions of Gödel's theorem. In particular, Wittgenstein deploys the idea that the meaning of a sentence is determined by its proof conditions. In virtue of the fact that there are object-level proofs and meta-level proofs (to put it in modern terminology), this still leaves the notions concerned in (I) ambiguous. Except for one circumstance, however, he thinks that once one clarifies the relevant meanings, the equivalence (I) should be rejected. In this case, no contradiction is forthcoming.

The one circumstance in which this is not the case is that in which *Principia* is inconsistent. In this case, he thinks, (I) is fine and contradiction arises. He also thinks that this does not pose any real problems. The thought that the inconsistency of *Principia* is unproblematic is not correct. Because *Principia* is based on an explosive logic, this means that all sentences would be provable, which renders it useless for most interesting purposes. To the extent that there are definite mistakes in the *Remarks*, this is plausibly one.

If we are dealing with an inconsistent metatheory, the matter is different, however. Provided that the metatheory is based on a paraconsistent logic, inconsistency may well be perfectly acceptable. Indeed, provided that we use an object-theory of arithmetic based on a paraconsistent logic, the same is the case. As I noted, such arithmetics, where both the Gödel sentence and its negation are provable – encoding 'Gödel's paradox' – are now well known. None of the formal material on paraconsistent logics and inconsistent arithmetics was, of course, known at the time Wittgenstein was writing. From an orthodox point of view, these possibilities could therefore have seemed wild. (Indeed, given that these techniques are still very unorthodox, the same might still be said.) But once one has taken these possibilities to heart, Wittgenstein's views on the countenancing of inconsistency are not at all wild, and the fact that he made them when he did shows striking prescience as well as groundbreaking originality. These are not the second-rate thoughts of an otherwise sparkling mind.

What one is to say about various of the positions taken by Wittgenstein, such as that mathematical sentences do not have propositional content, the redundancy account of truth, the acceptability paraconsistent logic, and so on, is, of course, another matter – and one well outside the brief of this article. But discussions of these matters can only be hindered by a misunderstanding of Wittgenstein's views. I hope that this article has substantially contributed to clearing such away.[20]

Notes

1 Some more recent commentators, notably Floyd, have been kinder about Wittgenstein's remarks. See Floyd (2001) for discussion and references.
2 Quotations are from Wittgenstein (1978: appendix III). All italics are original.
3 Note that Wittgenstein uses 'proposition' sometimes to mean 'indicative sentence', and sometimes to mean 'indicative sentence with propositional content'.
4 Strictly: '*P*' is not provable. But like Wittgenstein, I will suppress the quotation marks when no confusion can arise.
5 It must be admitted that the expression of the last sentence of 10 is a bit odd – though this is not a question of the translation, which is faithful. The only sense that the context would appear to admit is: 'But if this were now proved, or if I believed – perhaps through an error – that I had proved it, why should I not let the proof stand and [why] say I must withdraw my interpretation "*unprovable*"?'
6 It might be noted that the standard modern reaction would be to endorse the equivalence (I): *P* is not provable (in *Principia*) iff *P* (is true in the standard model of arithmetic). The arguments (A) and (B) stand (modulo the soundness of *Principia*): *P* is both true and unprovable. But this is no contradiction, since truth does not imply provability – Arithmetic is incomplete. But Wittgenstein, because he has, in effect, identified truth with provability, cannot draw this distinction.
7 See, e.g., p. 256 of the *Remarks*, and also Wittgenstein (1976: 209ff.).
8 See, e.g., (1976: 209).
9 See, e.g., Priest (1987: 59) and Priest (1995: 159).
10 For a comprehensive survey of paraconsistent logics, see Priest (2002). For a defence of the view that a paraconsistent logic is the correct logic to reason concerning paradoxes of self-reference, see Priest (1987).
11 There is one isolated remark – Wittgenstein (1976: 209) – that suggests that Wittgenstein countenanced the possibility that a contradiction entailed nothing – so endorsing some sort of connexive paraconsistent logic. But he never seems to have pursued this idea in detail.
12 Wittgenstein (*Remarks*: 322).
13 See, e.g., Priest (1997) and (2000).
14 On all this, see Priest (1994).
15 For a further discussion of Wittgenstein's verificationism/voluntarism, see Wright (1980) esp. pp. 364–86, who, following Dummett, calls the view radical conventionalsm.
16 Compare the case that Wittgenstein is considering with an orthodox understanding of the matter. According to this, Gödel gave us a metatheoretic proof of the unprovability of the undecidable sentence – on the assumption that *Principia* is sound. And the result of the proof really does show – on the same assumption – that there is no geometric construction of a certain kind. Since the proof predicate in *P* really does represent provability, (I) holds and so *P* is true in the standard model. Consistency is maintained, not by jettisoning (I), but by the distinction between proof and truth in the standard model.
17 Though one might note that an application of Wittgenstein's voluntarism might well be thought to get him off the hook here. Given a contradiction in *Principia*, one might simply *decide* not to apply the rule of Explosion.
18 A somewhat different interpretation of remark 17 was suggested to me by Brad Armour-Garb. According to this, Wittgenstein is enforcing the thought that different proofs of a sentence literally give it different senses – in the same way that operationalists say that the fact that there are different verifications of an empirical claim show it to be ambiguous. In this case, the fact that there are different proofs of the left and right sides of (I) demonstrates that it can be maintained only by equivocation.

Only if one and the same proof is the proof of both sides can (I) be maintained – in which case, a contradiction arises.

I think that the text does bear this interpretation, though I find it less plausible. For a start, though Wittgenstein is clearly tempted by the view that different proofs of the same theorem give it different senses, he does accept the fact that different proofs can have the same force (see, e.g., *Remarks*: 409). But even if he did accept the view in question, it is somewhat implausible to suppose that he is deploying it here. For if he were, he would hardly need to run through all the different cases and consider each in turn, as he does. He could just state the general argument, as I just have. At any rate, for the matter at hand, this is not an issue of great consequence: the upshot of this interpretation is exactly the same as the one I have given. We should give up (I), and so have consistency – unless *Principia* is itself inconsistent.

19 Note that there is a closing right-hand bracket missing in the text. The most plausible place to locate it would seem to be after 'Baroque style'.

20 Many thanks go to Bernhard Weiss for suggesting the topic of this essay, to the members of a discussion session at the University of St Andrews for helpful preliminary thoughts, to Stuart Candlish for advice on translation, and especially to Brad Armour-Garb and Bernhard Weiss for thoughtful comments on drafts of the paper.

References

Anderson, A. R. (1964) 'Mathematics and the "Language Game"' in P. Benacerraf and H. Putnam (eds) *Philosophy of Mathematics; Selected Readings*, Oxford: Blackwell, pp. 481–90 (first published 1958, *Review of Metaphysics* 11: 446–58).

Dummett, M. (1964) 'Wittgenstein's Philosophy of Mathematics' in P. Benacerraf and H. Putnam (eds) (1964) *Philosophy of Mathematics; Selected Readings*, Oxford: Blackwell, pp. 491–509 (first published 1959, *Philosophical Review* 68: 324–48).

Floyd, J. (2001) 'Prose *versus* Proof: Wittgenstein on Gödel, Tarski and Truth', *Philosophia Mathematica* 9: 280–307.

Kreisel, G. (1959) 'Wittgenstein's *Remarks on the Foundations of Mathematics*', *British Journal for the Philosophy of Science* 9: 135–58.

Priest, G. (1987) *In Contradiction*, Dordrecht: Kluwer Academic Publishers.

—— (1994) 'Is Arithmetic Consistent?' *Mind* 103: 337–49.

—— (1995) *Beyond the Limits of Thought*, Cambridge: Cambridge University Press; second (revised) edition 2002, Oxford: Oxford University Press.

—— (1997) 'Inconsistent Arithmetic, I: Finite Models', *Journal of Philosophical Logic* 26: 223–35.

—— (2000) 'Inconsistent Arithmetic, II: The General Case', *Journal of Symbolic Logic* 65: 1519–29.

—— (2002) 'Paraconsistent Logic', in D. Gabbay and F. Guenthner (eds) *Handbook of Philosophical Logic*, vol. 6 (2nd edition), Dordrecht: Kluwer Academic Publishers.

Wright, C. (1980) *Wittgenstein on the Foundations of Mathematics*, London: Duckworth.